

IX CONFERÊNCIA INTERNACIONAL SOBRE BIBLIOTECAS E REPOSITÓRIOS DIGITAIS DA AMÉRICA LATINA

BIREDIAL-ISTEC

SÃO PAULO • BRASIL

30 • 31 DE JULHO 1 • 2 DE AGOSTO 2019



Anais das sessões temáticas



IX CONFERÊNCIA INTERNACIONAL SOBRE BIBLIOTECAS E REPOSITÓRIOS DIGITAIS DA AMÉRICA LATINA BIREDIAL – ISTEC

Anais das sessões temáticas

São Paulo

2019

UNIVERSIDADE FEDERAL DE SÃO PAULO

Soraya Soubhi Smaili

Reitora

Nelson Sass

Vice-Reitor

Maria Eduarda dos Santos Puga

Diretora da Coordenadoria da Rede de Bibliotecas Unifesp (CRBU)

Norma Shizue Moriama Iwashita

Diretora Adjunta da Coordenadoria da Rede de Bibliotecas Unifesp (CRBU)

Conferência Internacional BIREDIAL-ISTEC (9. : 2019 : São Paulo, SP).

Anais das sessões temáticas. São Paulo: Universidade Federal de São Paulo, 2019.

ISBN: 978-85-62377-38-9

1. Ciência da Informação. 2. Biblioteconomia. 3. Biblioteca universitária. I. Universidade Federal de São Paulo. II. Instituto de Pesquisas Energéticas e Nucleares. III. Instituto Tecnológico de Aeronáutica. IV. Universidade Federal do ABC. V. Título.



Esta obra está licenciada com uma Licença [Creative Commons Atribuição-NãoComercial-Compartilha Igual 4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/).



ORGANIZAÇÃO E REALIZAÇÃO

Presidente

Maria Eduarda dos Santos Puga – Universidade Federal de São Paulo, Brasil

Comissão executiva

Andrea A. O. Hayashi - UNIFESP

Maria do Carmo Cardoso - UFABC

Cristiane M. Shirayama – UNIFESP

Vera Lucia P. R. Junqueira - ITA

Daianny S. de Oliveira - UNIFESP

Mery Piedad Z. Igami - IPEN

Diogo Misoguti - UNIFESP

Mayara R. Batista - UNIFESP

Flávio Castro de Sousa - UNIFESP

Richard W. Valdivia - UNIFESP

Juliana A. dos Santos - UNIFESP

Norma S. M. Iwashita - UNIFESP

ORGANIZAÇÃO E REALIZAÇÃO

Comissão Técnica-Científica

Maria Eduarda dos Santos Puga – Universidade Federal de São Paulo, Brasil

Caterina Groposo Pavão – Universidade Federal do Rio Grande do Sul, Brasil

Marisa Raquel De Giusti – ISTECS. Universidad Nacional de La Plata, Argentina

Malgorzata Lisowska – Universidad del Rosario, Colombia

Meilyn Garro – Universidad de Costa Rica, Costa Rica

Marlem Uribe Marengo – Universidad del Norte, Colombia

Pareceristas

Alexa Ramírez-Vega – ITCR

Andrea G. Nascimento – ICICT/FIOCRUZ

Ariel Jorge Lira - CESGI

Carlos Javier Nusch - CESGI

Caterina Groposo Pavão – UFRGS

Daniela Francescutti Martins Hott –
Câmara dos Deputados

Fabiano Couto Correa da Silva - FURG

Gonzalo Luján Villarreal – UNLP

Maria Helena F. Xavier da Silva – UFF

Marlem Uribe Marengo – UNINORTE

Mery Piedad Zamudio Igami – IPEN

Pablo de Albuquerque – UNLP

Rene Faustino Gabriel Junior – UFRGS

Rosalina Vazquez Tapia – UASLP

Samile Andréa de Souza Vanz – UFRGS

Susimery Vila Nova – UFPE

Thiago Magela R. Dias – CEFET/MG



Instituições sede



Organização



Patrocinadores



Local



SUMÁRIO

Apresentação.....	08
DIA 1 - REPOSITÓRIOS INSTITUCIONAIS E TEMÁTICOS.....	09
Verificação dos repositórios institucionais contemplados no projeto piloto do edital FINEP/IBICT/2009	10
Repositórios de acesso aberto brasileiros.....	24
Criação de um repositório temático no contexto das humanidades digitais para gestão da informação.....	37
Repositorio digital de microfotografías electrónicas: Una herramienta para la exploración y divulgación de micromundos ocultos.....	47
La vía verde del acceso abierto em Argentina: evolución y desarrollo de los repositorios en las universidades nacionales	61
DIA 2 - PUBLICAÇÕES ELETRÔNICAS, AVALIAÇÕES E MÉTRICAS	83
Avaliação ergonômica de repositórios digitais institucionais – revisão integrativa da literatura	84
Checklist de inspeção ergonômica de Repositórios Digitais Institucionais – CIERDI	101
Avaliação de confiabilidade do Repositório Institucional Arca	116
Rumbo a la autoevaluación: la elaboración de un Manual de Procedimientos de Digitalización.....	129
HUB-UR: Una solución de para el fortalecimiento de la visibilidad web de la Universidad del Rosario – Colombia.....	143
LUME - Módulo de estatísticas 2.0 para o Dspace 5.8	153
Integración de revistas desde repositorios institucionales hacia Open Journal Systems.....	164
A ciência e o livro eletrônico: o projeto BRAPCI Livros	177
Los datos crudos de investigación en las Humanidades – periódicos decimonónicos.....	186

A implantação da Base de Dados Científicos (BDC) da Universidade Federal do Paraná (UFPR)	195
Re-escritura da comunicação científica: um caminho para a ciência aberta.....	203
A multidimensionalidade da Ciência Aberta	218
Ciencias abierta y ciudadana: por la visibilidad de las mujeres en la investigación	236
Modelo de Visibilidade Acadêmica del Instituto Tecnológico de Costa Rica.....	250
DIA 3 - CIÊNCIA ABERTA, REPOSITÓRIO DE DADOS DE PESQUISA, PLANO DE GESTÃO DE DADOS	259
A integração do Arca - Repositório Institucional da Fiocruz com a Plataforma de Ciência de Dados aplicada à Saúde	260
Repositório bilíngue e a disseminação de conhecimento científico e cultural para e sobre educação de surdos	271
Desenvolvimento de Comunidades e Coleções por meio do Design Thinking: experiência do repositório do Instituto Nacional do Câncer José Alencar Gomes da Silva.....	282
Interoperabilidade de repositórios digitais: em busca da preservação digital	290
Preservação digital e agenda 2030 da IFLA	300
Sobre o projeto biblioteca do IPPUR 2.0	308
Acesso aberto, dados abertos e ciência aberta no Brasil: revisão sistemática de literatura a partir do Catálogo de Teses e Dissertações da CAPES	320
Visibilidade da ciência latinoamericana: contribuições da Rede LA Referencia para a ampliação do acesso aberto	333



Apresentação

Em sua nona edição, a *Conferência Internacional sobre Bibliotecas e Repositórios Digitais da América Latina* foi realizada na Universidade Nove de Julho, em São Paulo, Brasil, durante os dias 30 e 31 de julho e 01 e 02 de agosto de 2019 e foi organizada pela Universidade Federal de São Paulo (UNIFESP), Universidade Federal do ABC (UFABC), Instituto de Pesquisas Energéticas e Nucleares (IPEN) e Instituto Tecnológico de Aeronáutica (ITA), com a colaboração LibLink do Consórcio Ibero-Americano para a Educação em Ciência e Tecnologia (ISTEC).

A Conferência Internacional BIREDIAL-ISTEC dedica-se a compartilhar o conhecimento a uma ampla comunidade universitária e de diferentes disciplinas de conhecimento, comprometidas com a iniciativa de acesso aberto.

O objetivo do evento é expor, discutir e analisar as questões que envolvem a iniciativa: materiais para acesso aberto, o problema especial dos dados pesquisa, provisão, preservação e disseminação de obras, atenção aos direitos autorais e reprodutibilidade da investigação, capacitação dos recursos humanos para a gestão em plataformas digitais em acesso aberto.

Eixos temáticos gerais

REPOSITÓRIOS INSTITUCIONAIS E TEMÁTICOS: Gestão e Curadoria, Interoperabilidade, Visibilidade, Preservação digital e Avaliação de repositórios.

PUBLICAÇÕES ELETRÔNICAS, AVALIAÇÕES E MÉTRICAS: Impacto e Visibilidade na ciência, Preservação digital de periódicos eletrônicos.

CIÊNCIA ABERTA, REPOSITÓRIOS DE DADOS DE PESQUISA, PLANO DE GESTÃO DE DADOS: Práticas e políticas, Ética e Direitos autorais, Gestão e curadoria, Interoperabilidade, Visibilidade e promoção de dados abertos na pesquisa, Preservação digital e Reprodutibilidade da investigação na ciência.

A Conferência BIREDIAL-ISTEC é dirigida a Comunidades científicas, Observatórios, Bibliotecas, Centros de Dados Abertos e outros; Profissionais interessados em questões relacionadas à gestão, promoção e visibilidade do conhecimento; autoridades acadêmicas; Pesquisadores, Editores e Diretores de publicações científicas, alunos de várias especialidades; Instituições de Ensino Superior e Centros de Investigação; Pessoal do governo e da administração pública interessada em compartilhar preocupações relacionadas à transparência dos dados públicos.



DIA 1

REPOSITÓRIOS INSTITUCIONAIS E TEMÁTICOS

Gestão e curadoria

Interoperabilidade

Visibilidade

Preservação digital

Avaliação de repositórios

Verificação dos repositórios institucionais contemplados no projeto piloto do edital FINEP/IBICT/2009

Davilene Souza Santos¹; Flávia Goulart Mota Garcia Rosa²

¹ Bacharel em Biblioteconomia e Documentação - Universidade Federal da Bahia; Mestranda do Programa de Pós-Graduação Estudos Interdisciplinares Sobre a Universidade – EISU - Universidade Federal da Bahia. Email: davilenes13@gmail.com

² Doutora em Cultura e Sociedade; Mestra em Ciência da Informação - Universidade Federal da Bahia; Professora Associada IV - Universidade Federal da Bahia – UFBA. Email: fflaviagoulartroza@gmail.com

Resumo

Este estudo insere-se no tema sobre repositórios institucionais e temáticos, na avaliação de repositórios, e tem como proposta a verificação dos cinco repositórios institucionais contemplados com o kit tecnológico para sua implantação, através do edital da Financiadora de Estudos e Projetos e executado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia, como projeto piloto em 2009. As instituições que tiveram seus projetos aprovados para a implantação dos repositórios foram as seguintes: Universidade Federal da Bahia; Universidade Federal de Pernambuco; Universidade Federal de Santa Catarina; Universidade Federal do Rio Grande do Sul e a Universidade de Brasília. Essas instituições comprometeram-se em implantar os seus repositórios instituições com os recursos oferecidos pelo edital, e em contrapartida atuaram no desenvolvimento, manutenção e atualização destes, como forma de fomentar o acesso aberto na disponibilização de conteúdo. O objetivo deste trabalho é verificar o estágio atual de desenvolvimento e manutenção dos seus repositórios, e para isso foi realizada uma entrevista com os gestores, bem como a observação dos portais digitais desses, como instrumento de coleta de dados. A análise demonstra que alguns desses repositórios estão em um estágio mais avançado do que outros no que tange a versão do DSpace utilizada e sua disponibilização de conteúdo, bem como a gestão adotada. Concluímos que essa verificação tende a ser positiva no sentido de que compartilha as técnicas, os documentos e as práticas utilizadas pelos repositórios institucionais para vencer barreiras e se inserir no contexto do acesso aberto.

Palavras-chave: Comunicação científica. Movimento de Acesso Aberto. Repositórios Institucionais. Avaliação de repositórios institucionais.

Abstract

This study is part of the theme on institutional and thematic repositories, in the evaluation of repositories, and has as proposal the verification of the five institutional repositories contemplated with the technological kit for its implantation, through the edict of the Financier of Studies and Projects and executed by the Institute Brazilian Institute of Information in Science and Technology, as a pilot project in 2009. The institutions that had their projects approved for the implementation of the repositories were the following: Federal University of Bahia; Federal University of Pernambuco; Federal University of Santa Catarina; Federal University of Rio Grande do Sul and the University of Brasília. These institutions committed themselves to implanting their institutions' repositories with the resources offered by the public notice, and in return they would act in the development, maintenance and updating of these, as a way to foster open access in the provision of content. The objective of this work is to verify the current stage of development and maintenance of its repositories, for which an interview was conducted with the managers, as well as the observation of the digital portals of these, as a data collection instrument. The analysis shows that some of these repositories are at a more advanced stage than others in the version of DSpace used and its availability of content, as well as the management adopted. We conclude that this verification tends to be positive in that it shares the techniques, documents and practices used by institutional repositories to overcome barriers and to be inserted in the context of open access.

Keywords: Scientific communication. Movement of Open Access. Institutional repositories. Evaluation of institutional repositories.

Introdução

A comunicação científica, parte indispensável na produção do conhecimento, tem passado por diversas transformações ao longo do tempo. A forma como lidamos com os resultados de pesquisa tendem a ser o mais democrático possível no que tange a comunicação entre os pares e para a sociedade no geral. Dessa forma, o Movimento de Acesso Aberto (MAA) tem influenciado no compartilhamento das informações e permitido aos cientistas uma maior interação na disponibilização e consumo de conteúdo pelos usuários. No Brasil, o fomento ao MAA tem no Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) o condutor e incentivador das práticas do acesso aberto, atuando na busca da inserção do país no contexto mundial, já evidente em diversas outras nações.

Uma das iniciativas do IBICT após o lançamento do Manifesto de Apoio ao Acesso Aberto,¹ em 2005, foi o apoio e a execução do edital / FINEP/PCAL XBDB 002/2009² lançado pela Financiadora de Estudos e Projetos (FINEP), com o objetivo de contemplar instituições de ensino superior, e/ou centros de pesquisa com um kit tecnológico para implantação de seus repositórios institucionais, sendo esta, uma das propostas do MAA conhecida como Acesso Aberto Verde, no qual as instituições através dos seus repositórios, disponibilizam os conteúdos produzidos pela sua comunidade acadêmica e científica. Weitzel (2019) credita aos RI a possibilidade e a potencialidade para aumentar a visibilidade das produções científicas realizadas em países emergentes.

Nesse edital, além das mais de 20 instituições contempladas com o kit tecnológico, os projetos de cinco repositórios institucionais de Instituições Federais de Ensino Superior (IFES), foram selecionados como projeto piloto. As instituições foram: Universidade Federal da Bahia; Universidade Federal de Pernambuco; Universidade Federal de Santa Catarina; Universidade Federal do Rio Grande do Sul e a Universidade de Brasília. Estas, além do recebimento do kit tecnológico, receberam um servidor para implantação do RI, já com o software livre Dspace e sistemas operacionais instalados, bem como receberam assessoria para a implantação e gestão da ferramenta. A contrapartida para essas instituições seria justamente atuar no desenvolvimento, manutenção e atualização dos seus repositórios, fomentando assim os ideais do MAA.

O objetivo deste trabalho é verificar como esses cinco RI, que participaram do edital como projeto piloto, vêm atuando no desenvolvimento, manutenção e atualização dessa ferramenta, e como a sua utilização tem contribuído para fomentar o Acesso Aberto e se vêm, de fato, cumprindo com o papel que foi proposto.

Para essa análise, adotamos como coleta de dados a realização de entrevistas estruturadas, realizadas por meio eletrônico, com os gestores dos RI mencionados, com o intuito de identificarmos as barreiras enfrentadas e as práticas adotadas na implantação, desenvolvimento e manutenção dos repositórios, assim como a observação dos portais digitais dos repositórios, através da consulta sistemática aos sites dos RI, analisando as categorias selecionadas a fim de verificarmos a versão do Dspace utilizada, disponibilização de conteúdo, tipologia dos documentos, estrutura e interação. Para tal foi criado um instrumento denominado "Roteiro de observação e análise do RI" do qual consta as categorias de observação.

¹ Disponível em: <http://livroaberto.ibict.br/Manifesto.pdf>. Acesso em: 10 abr. 2019.

² Disponível em: <https://kuramoto.wordpress.com/2009/04/20/edital-fineppcalxbdb-n%C2%BA-0012009-foi-revogado/>. Acesso em: 10 abr. 2019.

Avaliação de Repositórios Institucionais

Para pontuarmos a avaliação de Repositórios Institucionais (RI), é primordial que possamos definir o que venha a ser o RI.

Segundo Bustos-Gonzalez; Fernandez-Porcel e Johnson (2007), na obra, *Diretrizes para a criação dos repositórios institucionais nas universidades e organizações de educação superior*, RI é:

Um sistema de informação que reúne, preserva, divulga e permite o acesso à produção intelectual e acadêmica das Comunidades Universitárias. Na atualidade o RI representa uma ferramenta chave no sistema político científico e acadêmico da universidade.

Corroborando com os autores citados, Nascimento (2014, p. 23) considera que “Os repositórios digitais se constituem como ferramentas tecnológicas de informação e comunicação favorecendo a disseminação da produção do conhecimento”.

No que tange a avaliação dos repositórios institucionais, esta é parte essencial no processo da comunicação científica em acesso aberto, que tem por objetivo um acompanhamento periódico da ferramenta, em virtude da sua recente concepção, como apontado por Vianna e Carvalho (2013, p. 13)

Os RIs são ferramentas relativamente novas, em desenvolvimento e em constante mudança e a percepção de seus resultados positivos ainda é pequeno por parte dos usuários. Como ferramenta em desenvolvimento ela necessita de avaliações constantes e uma integração entre os profissionais da área da Ciência da Informação e o pessoal de TI, bem como com os seus usuários.

Alguns fatores são essenciais para o sucesso e amplitude de um RI, dentre eles, a avaliação contínua, objetivando identificar pontos frágeis que necessitem de uma intervenção, bem como pontos fortes que possam ser ampliados, socializada e compartilhada com comunidades afins.

Lameira (2016), ao propor um método de avaliação para os RI, nos apresenta algumas características importantes para o sucesso de um repositório institucional. Essas características foram identificadas com base em diversos autores, listados no seu estudo. A autora nos apresenta três dimensões de avaliação, são elas: qualidade do sistema; qualidade da informação e qualidade do serviço. Para cada dimensão apresentada existem diversas características que podem ser avaliadas. Para exemplificar, seguem algumas delas no quadro abaixo:

Quadro 1: Dimensões e características de avaliação do RI.

Dimensões	Características						
Qualidade do Sistema	Utilidade	Facilidade de uso	Confiabilidade	Tempo de resposta	Integração	Flexibilidade	Acessibilidade
Qualidade da Informação	Clareza	Relevância	Legibilidade	Utilidade	Exatidão	Atualização	Oportunidade
Qualidade do Serviço	Credibilidade		Confiabilidade	Comunicação		Segurança	Acessibilidade

Fonte: elaboração do autor.

Assim como nesse levantamento, Lameira (2016) sugere que o método de observação dos portais, bem como entrevistas com os gestores dos RI, sejam os mais eficazes para a obtenção das respostas a cada uma dessas características, a fim de se realizar uma avaliação que corresponda de fato a situação real.

Uma das características referente a qualidade do serviço trata-se da credibilidade, e esta pode ser conferida através da posição ocupada pelo RI no “*webometrics*”,³ que consiste em um sistema de classificação de universidades em âmbito mundial criado em 2008, que avalia, dentre outros aspectos, a quantidade de conteúdo da instituição disponível na rede mundial de computadores, a visibilidade e o impacto dessas publicações e suas citações, gerando um *Ranking* das universidades. De acordo com Santos e Cardoso (2013, p. 02) “O *Ranking* tornou-se, com o passar dos anos, uma referência básica para perceber a visibilidade de repositórios científicos na web”.

Os repositórios analisados nesse estudo aparecem nas seguintes posições desse *Ranking*:

Quadro 2: Ranking webometrics

Universidade	UFBA	UFPE	UnB	UFRGS	UFSC
Ranking	12 ^a	16 ^a	9 ^a	5 ^a	7 ^a

Fonte: https://www.webometrics.info/en/Latin_America/Brazil.

Nessa análise podemos verificar que a Universidade Federal da Bahia está a frente apenas da Universidade Federal de Pernambuco em termos de disponibilização de conteúdo na WEB, de acordo com o “*Ranking webometrics*”, levando-se em consideração as cinco universidades objeto desse estudo.

³ Disponível em: https://www.webometrics.info/en/Latin_America/Brazil. Acesso em: 05 maio 2019.

De um modo geral, percebemos que os cinco repositórios em questão possuem boas colocações nesse *Ranking*. Destacamos, no entanto, que as universidades que estão na região Nordeste do Brasil ainda possuem uma colocação inferior aos RI localizados no Sul e Centro-Oeste do País.

O RI ao executar uma avaliação de forma periódica, possibilita que algumas falhas possam ser detectadas e prontamente corrigidas. Além do mais, nesse processo de autoavaliação também pode ser encontrado fatores positivos passíveis de socialização entre os demais RI, constituindo assim uma dinâmica de compartilhamento em prol do melhoramento dessa estratégia prevista no MAA. É nesse sentido que se justifica a necessidade de verificação, juntos aos cinco repositórios selecionados pelo edital supramencionado, do cenário atual de desenvolvimento, manutenção e disponibilização de conteúdo.

No que se refere a autoavaliação, verificamos que o RI-UFSC apresentou recentemente um “Relatório de avaliação e diagnóstico do repositório institucional da Universidade Federal de Santa Catarina” (2017), que possibilita uma análise acerca da implantação, manutenção e gestão da ferramenta.

Entrevista com os gestores do RI

Para a verificação de algumas características nos RI existe a necessidade de extração desses dados através dos gestores dos repositórios, conforme aponta Lameira (2016), para tanto, um dos instrumentos de coleta de dados utilizado para identificar as barreiras enfrentadas e as práticas adotadas na implantação, desenvolvimento e manutenção dos repositórios, foi a realização de uma entrevista com os gestores que fazem parte do universo da pesquisa.

Essas entrevistas estruturadas foram realizadas através do envio de e-mail aos endereços eletrônicos, em 25 de julho de 2018, tanto dos repositórios quanto dos próprios gestores, quando havia esse contato. Obtivemos resposta de quatro dos cinco gestores. O gestor que não respondeu foi contactado inúmeras vezes, sem sucesso.

As perguntas dirigidas aos gestores, de forma individualizada, nos possibilitaram extrair deles a sua percepção como participante no projeto piloto do edital, bem como, o modo de atuação dentro de uma esfera institucional tão ampla quanto as Universidades Federais no Brasil. Foram realizadas três perguntas:

1. O Repositório Institucional da Universidade (Nome da instituição) fez parte do projeto piloto do IBICT a partir do edital FINEP/PCAL/XBDB 002 / 2009, quais os pontos positivos dessa participação? Houve pontos negativos?
2. Quais as principais dificuldades encontradas para a implantação do RI?
3. Como avalia o RI dessa instituição após nove anos do edital?

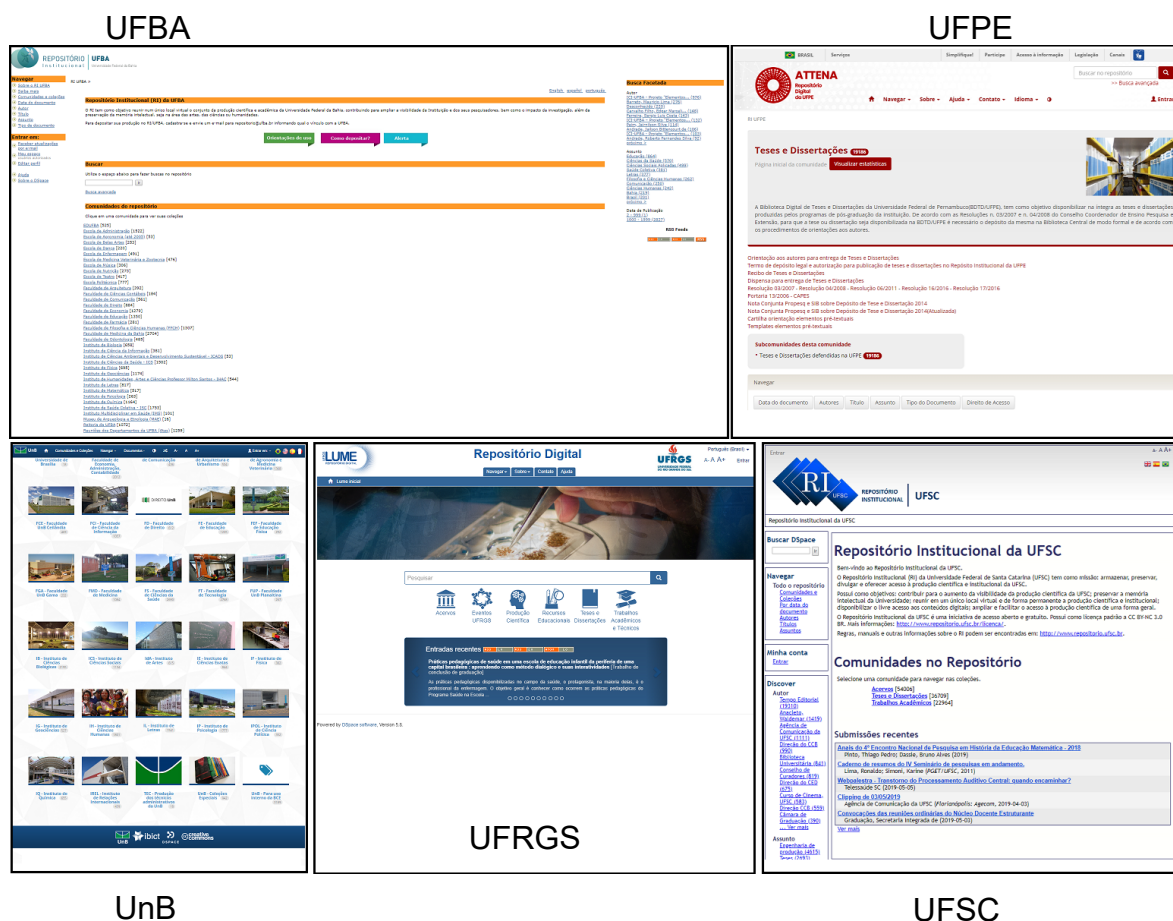
Observação dos portais digitais dos Repositórios Institucionais

Para a análise dos RI foi realizada a observação dos próprios repositórios institucionais, tendo como categorias de verificação os seguintes pontos:

1. Layout da página;
2. Disposição do conteúdo da página principal;
3. Versão DSpace;
4. Disponibilização da versão em outro idioma, e quais;
5. Existência de documentos normativos, e quais;
6. Tipos de documentos disponibilizados;
7. Quantidade de documentos disponibilizados;

Aqui, dispomos na Figura 1 um *print* com o *layout* das páginas dos cinco RI e a disposição de conteúdo de forma simplificada.

Figura 1: Página inicial dos RI



UnB

UFSC

Fonte: elaboração do autor a partir das páginas dos RI universo da pesquisa.

Através dessa disposição é possível identificar que, apenas ao ter o primeiro contato com o RI, já percebemos a diferença entre as versões apresentadas do DSpace, software utilizado na implantação do RI, e que

consequentemente irá impactar o usuário. Alguns RI apresentam imagens na primeira página, o que certamente é possibilitado pela versão atualizada do software, e contribui para impactar positivamente, no usuário, através de uma linguagem visual.

Esse fator nos remete a segunda categoria de análise, ou seja, a disponibilização do conteúdo na página principal. Ao analisarmos as imagens, podemos perceber que um sistema de identificação que explora elementos visuais influenciará na busca da área do conhecimento que se deseja consultar, ou mesmo das comunidades.

Vemos aqui que a versão atualizada do software DSpace é um fator que causa um forte impacto no visual e consequentemente no acesso ao RI, portanto, é importante que os gestores e as instituições mantenedoras estejam atentas quanto a atualização não somente do conteúdo nos Repositórios, mas também da versão do software, visto que este poderá expandir as possibilidades em termos visuais, de uso e de interação.

Como já citado a versão do DSpace possui sua parcela de contribuição na visibilidade do RI, logo, procuramos identificar as versões que os cinco RI estão utilizando. No Quadro 3, relacionamos as versões encontradas na página principal dos RI, em dois deles esse dado não aparece.

Quadro 3: Categorias analisadas nos repositórios

	UFBA	UFPE	UnB	UFRGS	UFSC
Versão DSpace	3.2 Em fase de atualização para a versão 5.7	5.6	-	5.8	-
Idiomas	Português Espanhol Inglês	Português Espanhol Inglês	Português Espanhol Inglês Francês	Português Espanhol Inglês	Português Espanhol Inglês

Fonte: elaboração do autor.

Outro fator que pode gerar um impacto positivo é a possibilidade da página do RI ser pesquisada em outro idioma. Nesse sentido, verificamos que os cinco RI verificados possuem ao menos três idiomas disponíveis, destacando-se o RI da UnB que possuem além dos idiomas Português, Inglês e Espanhol, possui também o idioma Francês.

No que se refere aos documentos normativos disponíveis na página dos RI, relacionamos abaixo alguns desses documentos encontrados.

Quadro 4: Documentos normativos disponíveis no portal dos RI

UFBA	UFPE	UnB	UFRGS	UFSC
Em Política Institucional (Consta a <i>Portaria de criação do RI</i>) (2010).	Resolução que institui a política de informação do RI (2017).	Portaria que Institui a divulgação digital das teses e dissertações.	Portaria que institui a política de informação para o Lume (2010)	Portaria que sistematiza a disponibilização (2013)
Tutorial para submissões no repositório institucional (RI) da Universidade Federal da Bahia (UFBA).	Portaria de criação do grupo gestor (2015).	Resolução da Reitoria (2013) *Regulamenta a Política de Informação do Repositório Institucional da Universidade de Brasília (RIUnB).	Instrução Normativa (2010) Todo Trabalho de Conclusão de Curso de Graduação deverá ser disponibilizado no RI.	Relatório de avaliação e diagnóstico do repositório institucional da universidade federal de santa Catarina (2017).
Orientações para uso do Repositório Institucional da UFBA.	Portaria de criação do grupo gestor (2019).		Portaria que institui o Comitê gestor (2010).	Portaria para sistematizar dos TCC no RI-UFSC.
	Legislação pertinente ao acesso a informação (Leis).		Resolução (2005) – Estabelece a entrega da versão impressa e digital das teses e dissertações.	Portaria MEC (2006) - Institui a divulgação digital das teses e dissertações produzidas pelos programas de doutorado e mestrado reconhecidos.
			Instruções para submissão de teses, dissertações, trabalhos de conclusão de mestrado e de especialização para divulgação no lume – repositório digital da UFRGS.	

Fonte: elaboração do autor.

Observamos que apenas no RI da UFSC os documentos normativos são obtidos através de um link que remete para outra página, diferente da página original do RI. A página encontra-se organizada, porém o direcionamento para outro ambiente pode ser um fator de impedimento e desconhecimento ao usuário.

Os documentos normativos encontrados não contemplam a recomendação do OpenDoar⁴ - Diretório de Repositórios de Acesso Aberto, que propõe a disponibilização de algumas políticas que abarquem a submissão, os dados e a preservação nos RI. Esse fato já havia sido identificado por Silva Júnior (2017), que ao analisar as IFES brasileiras que possuem RI, verificou que estas não atendiam a uma dessas recomendações do OpenDoar, a divulgação de uma política de preservação.

Os tipos de documentos disponibilizados nos RI, de uma forma geral, são: teses e dissertações; artigos em eventos; artigos em periódicos; capítulos de livros e livros. Nem todos os Repositórios contemplam essas categorias de documentos, mas de forma ampla, todos eles disponibilizam teses e dissertações.

Quadro 5: Tipos de documentos disponíveis nos Repositórios

	Teses e Dissertações	Artigos em Eventos	Artigos em Periódicos	Capítulos de Livros	Livros
UFBA	Contém	Contém	Contém	Contém	Contém
UFPE	Contém	Não localizado	Não localizado	Não localizado	Não localizado
UnB	Contém	Contém	Contém	Contém	Contém
UFRGS	Contém	Contém	Contém	Contém	Contém
UFSC	Contém	Não localizado	Contém	Não localizado	Contém

Fonte: elaboração do autor.

Disseminação do conteúdo disponibilizado nos RI

Destacamos que o Repositório da Universidade Federal da Bahia possui um serviço de disseminação do conhecimento denominado “Alerta”,⁵ que se destina a

⁴ Disponível em: <http://v2.sherpa.ac.uk/opensoar/>. Acesso em: 10 de maio 2019.

⁵ Disponível em: <https://ndc.ufba.br/>. Acesso em: 09 maio 2019.

divulgar a produção acadêmica da instituição disponível nesse ambiente virtual. Essa divulgação através de um serviço disseminação contribui para ampliar a abrangência e a visibilidade da produção científica e acadêmica produzida na instituição, com envios de Boletins em formato de newsletter semanais a toda comunidade da UFBA, de parte da produção científica e acadêmica inserida no RI recentemente.

Um serviço dessa natureza não foi identificado nos demais repositórios objeto desse estudo, dessa forma, esse aspecto positivo do RI-UFBA pode influenciar os outros participantes do projeto piloto a ampliarem o seu raio de divulgação, através da criação de um serviço semelhante.

Análises e discussões

Através das respostas obtidas, a partir das entrevistas, foi possível verificar o importante papel desempenhado pelo IBICT na execução do edital, bem como as dificuldades enfrentadas por alguns desses repositórios, pontos comuns aparecem nas respostas de quase todos os gestores, a exemplo da ausência de uma política institucional para o RI. Esse fator foi identificado através da constatação da ausência em alguns desses RI de uma política de fato institucional que pudesse fortalecer o MAA, como demonstrado no Quadro 4.

A troca de experiências entre as instituições foi considerada um fator positivo, inclusive no que se referia a utilização do DSpace. O RI – UFRGS relata o quão importante foi a participação no projeto piloto, mesmo já tendo ocorrido a implantação do “Lume” – nome próprio associado ao RI da instituição, em 2008, a troca de experiência, possibilitou contribuir com as outras instituições no que se referia ao uso do Dspace.

Consideramos que a troca de experiência entre as instituições pode ter contribuído também na implantação do Repositório Instituição da UFPE em 2014, e que, recentemente, em 2019 passou a adotar o nome próprio de “Antena”. Verificamos que a adoção de nome próprio dado aos repositórios cria uma identidade forte e um significado importante ao sugerir nomes ligados ao conhecimento. Tratando-se dos cinco RI verificados, apenas as Universidade do Rio Grande do Sul e Pernambuco adotam uma denominação própria.

Identificamos que algumas instituições possuem uma atenção mais aprimorada no que tange a gestão do RI, demonstrando interesse em executar de forma adequada as diretrizes do MAA, mas podem esbarrar em entraves que na maioria das vezes independe da vontade e empenho do gestor e da equipe do RI.

No geral, a autoavaliação dos RI revela os avanços alcançados em termos de comunicação científica da instituição e visibilidade dessa produção, ao mesmo tempo em que traduz a consciência da necessidade de maior participação da comunidade e das instâncias superiores.

Em relação à observação realizada nos repositórios, foi possível verificar que em alguns deles, encontram-se em suas versões mais atualizadas do Dspace, sendo que um deles a efetivação dessa atualização está em processo, o que

certamente possibilitará uma melhor visualização do conteúdo com um layout mais moderno, tornando-o mais agradável e dinâmico, e outros dois que não foi possível identificar a versão.

Outro aspecto relevante diz respeito ao idioma, todos eles possuem pelo menos a versão em três idiomas, a saber: Português, Inglês e Espanhol – facilitando dessa forma o intercâmbio de informações. A exceção fica acerca do RI da UnB que possui adicionalmente o idioma Francês.

Considerações Finais

O estudo nos permite concluir que a implantação dos RI nas IFES selecionadas para o projeto piloto, contribuiu de forma significativa para ampliar a comunicação científica da instituição, mas ainda enfrenta dificuldades que permanecem desde a implantação, tais como: participação da comunidade acadêmica, engajamento das instâncias superiores e recursos humanos para manutenção. Foi possível identificar que alguns RI estão mais avançados em termos de disponibilização de conteúdo, como em apresentação e layout, trazendo talvez uma maior satisfação ao usuário.

Identificamos que alguns desses repositórios ainda não dispõem de documentos normativos adequados, especialmente a política institucional bem definida, o que evidencia um fator negativo, haja vista esse ser o documento norteador para a comunidade acadêmica atuar na participação mais efetiva na disponibilização do conteúdo produzido pela universidade.

Considerando que as IFES selecionadas como projeto piloto no edital estavam em estágios diferentes na concepção do acesso aberto, e que de fato todas elas implantaram os seus RI com o apoio do IBICT, ainda que com certa dificuldade de ordem institucional, verificamos que estas têm cumprido os objetivos do MAA, algumas delas ainda de forma tímida e incipiente. A avaliação constante do estado de desenvolvimento e manutenção dos RI é uma premissa que deve ser motivada e executada com a finalidade de aprimoramento e acompanhamento das tendências em torno do *Open Access*.

Referências

BARBOSA, Gesmar Kingeski (Coord.). **Relatório de avaliação e diagnóstico do repositório institucional da Universidade Federal de Santa Catarina**. UFSC: Santa Catarina, 2017.

BUSTOS-GONZALEZ, Atílio; FERNANDEZ-PORCEL, Antonio; JOHNSON, Ian M. **Diretrizes para criação dos repositórios institucionais nas universidades e organizações de educação superior**. Babel Library: Alfa Network Babel Library. 2007. Disponível em:

http://eprints.rclis.org/13512/3/Directrizes_RI_portugues.pdf. Acesso em: 07 mai. 2019.

LAMEIRA, Ana Kelly Alves. Avaliação de repositórios institucionais brasileiros: uma proposta de método de avaliação. **Cadernos BAD**, Lisboa, n. 2, p. 153-167, 2016. Disponível em: <http://www.brapci.inf.br/index.php/res/v/61863>. Acesso em: 07 mai. 2019.

NASCIMENTO, Aline Vieira do. **Repositórios digitais**: identificando fatores de sucesso para as bibliotecas digitais e repositórios institucionais. 2014. 113 f.; Dissertação (Mestrado em Ciência da Informação) - Escola de Comunicação - Universidade Federal do Rio de Janeiro / Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2014. Disponível em: http://ridi.ibict.br/bitstream/123456789/892/1/DISSERTA%C3%87%C3%83O_ALINENASCIMENTO_MAR2013.pdf. Acesso em: 15 out. 2018.

SANTOS, Rafael Antunes dos; CARDOSO, Roberto Carlos. Avaliação de repositórios institucionais: o Brasil no ranking webométrico. *In*: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO. 25, 2013, Florianópolis. **Anais [...]**, Florianópolis: FEBAB, 2013. Disponível em: <https://portal.febab.org.br/anais/article/view/1387>. Acesso em: 07 mai. 2019.

SILVA JR, Laerte Pereira da. **Os Repositórios Institucionais da Universidades Federais do Brasil**: um modelo de política de preservação digital. 2017. 190 f. Tese (Doutoramento em Informação e Comunicação em Plataformas Digitais). Faculdade de Letras, Universidade do Porto, Porto, 2017. Disponível em: <https://repositorio-aberto.up.pt/bitstream/10216/105842/2/202343.pdf>. Acesso em: 12 out. 18

VIANNA, Sheila Maria de Vasconcellos; CARVALHO, Rogério Atem de. Benefícios da implantação de repositórios institucionais na preservação da memória institucional. *In*: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO. 25, 2013, Florianópolis. **Anais[...]**, Florianópolis: FEBAB, 2013. p. 1-15. Disponível em: <<https://portal.febab.org.br/anais/article/view/1613/1614>>. Acesso em: 15 dez. 2017.

WEITZEL, Simone da Rocha. O mapeamento dos repositórios institucionais brasileiros: perfil e desafios. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Niterói, v. 24, n. 54, p. 105-123, jan./abr. 2019. Disponível em:



<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2019v24n54p105/38092>. Acesso em: 11 abr. 2019.

Repositórios de acesso aberto brasileiros: características, crescimento e possibilidades futuras

Raphael Figueiredo Xavier¹

¹ Bibliotecário. Fundação Getúlio Vargas – FGV. Email: raphael.xavier@fgv.br

Resumo

Este trabalho, adequado ao eixo temático "Repositórios Institucionais e Temáticos", busca avaliar a evolução de repositórios brasileiros entre 2011 e início de 2019 e com isso proporcionar uma visão global da infraestrutura destas plataformas. O surgimento de repositórios está diretamente ligado a popularização de tecnologias de informação e comunicação e seu impacto no processo da comunicação científica. Com o avanço do movimento de políticas de acesso aberto em todo o mundo, os repositórios tiveram um desenvolvimento relevante e significativo na última década. Através do diretório OpenDOAR, pretende-se verificar a quantidade de repositórios brasileiros, o idioma do conteúdo, visão geral dos softwares utilizados nas plataformas, assunto e conteúdo das coleções e o crescimento no período de 2011 a 2019. Com isso, é possível considerarmos trajetórias futuras no desenvolvimento de repositórios como a possibilidade de infraestrutura distribuída globalmente e serviços de valor agregado, transformando o sistema mais centrado em pesquisa, aberto e apoiando a inovação, enquanto, ao mesmo tempo, é coletivamente gerenciado pela comunidade acadêmica.

Palavras-chave: Repositórios Digitais; Open-Access; Movimento de acesso livre a informação; OpenDOAR

Abstract

This work, appropriate to the thematic axis "Institutional and Thematic Repositories", seeks to evaluate the evolution of Brazilian repositories between 2011 and early 2019 and thus provide a global view of the infrastructure of these platforms. The emergence of repositories is directly linked to the popularization of information and communication technologies and their impact on the process of scientific communication. With the advancement of open access policies worldwide, repositories have developed relevant and significantly over the last decade. Through the OpenDoar directory, we intend to check the number of Brazilian repositories, content language, an overview of the software used in these platforms, subject and

content of the collections and growth in the period from 2011 to 2019. With this, it is possible to consider possible future paths in the development of repositories such as the possibility of globally distributed infrastructure and value-added services, transforming the system more focused on research, open and supported innovation, while at the same time collectively managed by the academic community.

Keywords: Digital repositories; Open-Access; OpenDOAR

Introdução

Repositórios institucionais podem ser descritos como coleções digitais que coletam e preservam a produção intelectual de uma comunidade específica (PINFIELD, 2009). Estes repositórios armazenam e tornam acessíveis o patrimônio educacional, de pesquisa e associado a uma instituição. O surgimento destas coleções está diretamente atrelado ao desenvolvimento de tecnologias de informação e comunicação e seu impacto no processo de comunicação científica. Com o avanço do movimento de políticas de acesso aberto em todo mundo (Baseada na declaração de Budapeste, em 2001), os repositórios tiveram um desenvolvimento relevante e significativo na última década.

Diante deste fato, esta pesquisa busca avaliar a evolução de Repositórios brasileiros entre 2011 e início de 2019 e com isso proporcionar uma visão holística da infraestrutura destas plataformas. Os dados para esta pesquisa são de origem do OpenDOAR, um diretório que reúne iniciativas de acesso aberto e monitoram o crescimento do movimento de acesso aberto à informação científica no mundo através do registro de repositórios acadêmicos Open-Access. Nesse sentido, implica, por um lado, abordagem teórica sobre a temática acerca dos conceitos basilares e, de outro, abordagem exploratória e comparativa no que tange o objeto empírico: os repositórios brasileiros registrados no OpenDOAR.

Com isto, além de delinear uma progressão histórica e as características atuais da situação dos repositórios, a pesquisa traz discussões acerca de possíveis trajetórias futuras no desenvolvimento de repositórios como a possibilidade de infraestrutura distribuída globalmente e serviços de valor agregado, transformando o sistema mais centrado em pesquisa, aberto e apoiando a inovação e, ao mesmo tempo, coletivamente gerenciado pela comunidade acadêmica.

Definições

Inicialmente, é importante estabelecer algumas definições de conceitos-chave, particularmente 'Acesso aberto a Informação' e 'Repositórios', por serem fundamentos bases para esta pesquisa.

O Acesso aberto a informação é respectivamente o resultado:

- (1) de uma reação dos pesquisadores ao modelo de negócios de editoras comerciais de revistas científicas (e seus preços cada vez mais altos de

assinatura); e da (2) crescente conscientização do aumento de impacto provocado pela disponibilização de documentos científicos livres de barreiras ao acesso. O mote do movimento mundial em favor do Acesso Livre a resultados de pesquisa, portanto, é a disseminação ampla e irrestrita dos resultados de pesquisas financiadas com recursos públicos (BAPTISTA et al, 2007, p.2).

No início do século XXI a Open Society Institute (OSI) promoveu uma reunião em Budapeste que resultou, provavelmente em um dos mais relevantes documentos no marco das iniciativas do movimento de Acesso Livre. O documento ficou conhecido como Budapest Open Access Initiative (BOAI)¹. A BOAI estabeleceu ali duas estratégias práticas para iniciativa Open-Access: uma seria através de revistas de acesso livre, sem restrição de acesso e sem cobrança de assinatura ou taxas de acesso; outra seria pelo auto arquivamento, através dos autores, de artigos publicados nas revistas científicas em repositórios, disciplinares ou institucionais. A primeira foi denominada de via dourada e a segunda, via verde. (HOUGHTON; SWAN, 2013)

Ambos mecanismos possuem larga literatura discutindo suas especificidades e efeitos. Neste ano, o Directory of Open Access Journals – DOAJ registrou 12.932 revistas de Acesso livre que cumprem diretrizes da via dourada², enquanto o OpenDOAR tem inscritos em março de 2019 cerca de 4.000 Repositórios Institucionais operando na via verde³. Pinfield (2009) nos recorda que esses dois caminhos são vias paralelas distintas, alternativas e não complementares e que defensores de ambos os lados tem apresentado discussões na literatura da área.

O que é perceptível é que a discussão acerca de repositórios está vinculada diretamente ao seu papel de suporte à Via Verde (PINFIELD, 2009; BJÖRK ET AL., 2014; HOUGHTON e SWAN, 2013) e os Repositórios incluídos neste estudo estão neste escopo. Pinfield (2009, p.165, tradução nossa) define os Repositórios como:

Um conjunto de sistemas e serviços que facilita a ingestão, o armazenamento, o gerenciamento, a recuperação, a disponibilização e a reutilização de objetos digitais. Repositórios podem ser criados por instituições, comunidades, financiadores de pesquisa ou outros grupos. Podem fornecer acesso a uma variedade de objetos digitais, incluindo artigos de periódicos avaliado por pares, capítulos de livros, teses, conjuntos de dados, objetos de aprendizado ou arquivos de mídia. (TRADUÇÃO NOSSA)⁴

¹ A declaração original pode ser lida em: <https://www.budapestopenaccessinitiative.org/read>

² Acessar a página oficial do DOAJ. Disponível em <https://doaj.org/>. Acesso em: 01/03/2019

³ Acessar a página oficial do OpenDOAR. Disponível em <https://v2.sherpa.ac.uk/pendoar/>. Acesso em: 01/03/2019

⁴ As a set of systems and services that facilitates the ingest, storage, management, retrieval, display, and reuse of digital objects. Repositories may be set up by institutions, subject communities, research funders, or other groups. They may provide access to a variety of digital objects, including peer-reviewed journal articles, book chapters, theses, datasets, learning objects, or rich media files.

Distinções, tipologias e características são muito bem colocadas em Heery e Anderson (2005). Aqui, cabe destacar que o presente artigo analisa toda população de repositórios brasileiros de acesso aberto através do OpenDOAR, um diretório de repositórios de acesso aberto baseado na Universidade de Nottingham, Reino Unido, realizado inicialmente em parceria com a Universidade de Lund, na Suécia. O OpenDOAR foi criado em 2005 para criar “um diretório oficial de repositórios acadêmicos de acesso aberto” em todo o mundo (OPENDOAR). Seus registros são periodicamente revisados e atualizados em seu banco de dados oferecendo um escopo atualizado do cenário de repositórios em todo mundo.

O serviço OpenDOAR é, atualmente, usado como fonte de dados para pesquisas de acesso aberto e diferentes projetos de monitoramento dos avanços do Open Access (projetos como o BASE, CORE, Open Access Tracker, OpenAIRE, OpenDepot, Repository 66 e ROAR).

Resultados

Crescimento dos repositórios

No geral, o número total de repositórios brasileiros no OpenDOAR cresceu de 28 em janeiro de 2011 para 109 em abril de 2019 (conforme mostra a figura 1). Isso representa um aumento de 289% durante o período. No entanto, a taxa de crescimento não foi inteiramente consistente e equilibrada durante esse período: houve crescimento muito acentuado em 2011 (de 28 para 56 repositórios entre janeiro e dezembro) e um período estagnado de crescimento em todo o ano 2018.

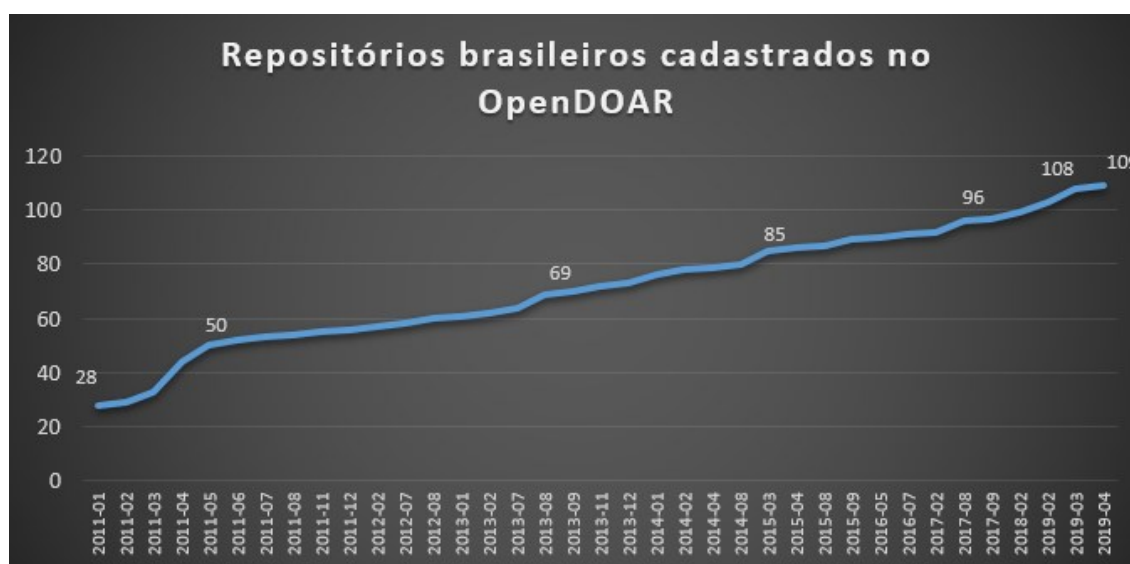


Figura 1: Evolução de repositórios brasileiros cadastrados no OpenDOAR

Distribuição geográfica

A Figura 2 ilustra o crescimento de repositórios digitais no Brasil divididos por região geográfica entre 2011 e abril de 2019.

Os dados mostram a região sudeste com maior consistência diante as outras regiões, que por sua vez tiveram leves crescimentos em meio a longos períodos estagnados. A Região Norte registrou seu primeiro Repositório somente em 2018.

Mais detalhes podem ser apreciados na Figura 3, onde o detalhe de crescimento é mostrado pelas Unidades Federativas do país. Os dados mostram que o estado de São Paulo teve a maior cobertura no OpenDOAR, seguido pelo Distrito Federal, Rio de Janeiro e Minas Gerais. Aqui, podemos perceber que apesar da região Centro-Oeste parecer como segunda maior região coberta, seu valor se deve quase que exclusivamente ao Distrito Federal, que detém a marca de 18 dos 22 repositórios registrados na região. Voltando a observação dos estados, percebemos um segundo bloco que agrupa os 3 estados da Região Sul: Paraná, Rio Grande do Sul e Santa Catarina e, por fim, um último grupo que mostra estados com 2 ou apenas 1 repositório registrado.

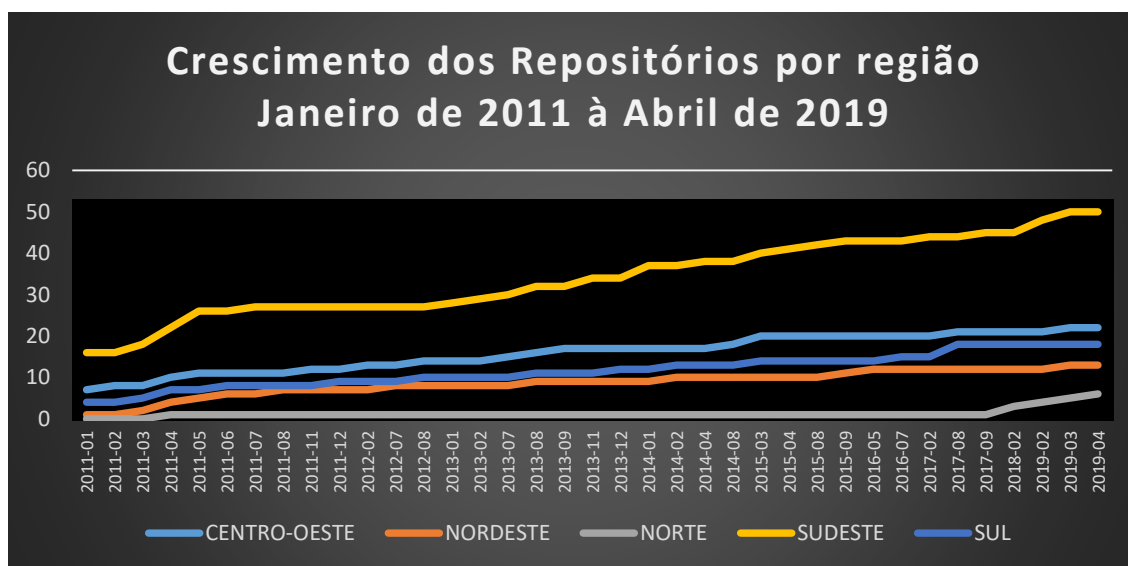


Figura 2: Crescimento de repositórios brasileiros por região geográfica no OpenDOAR

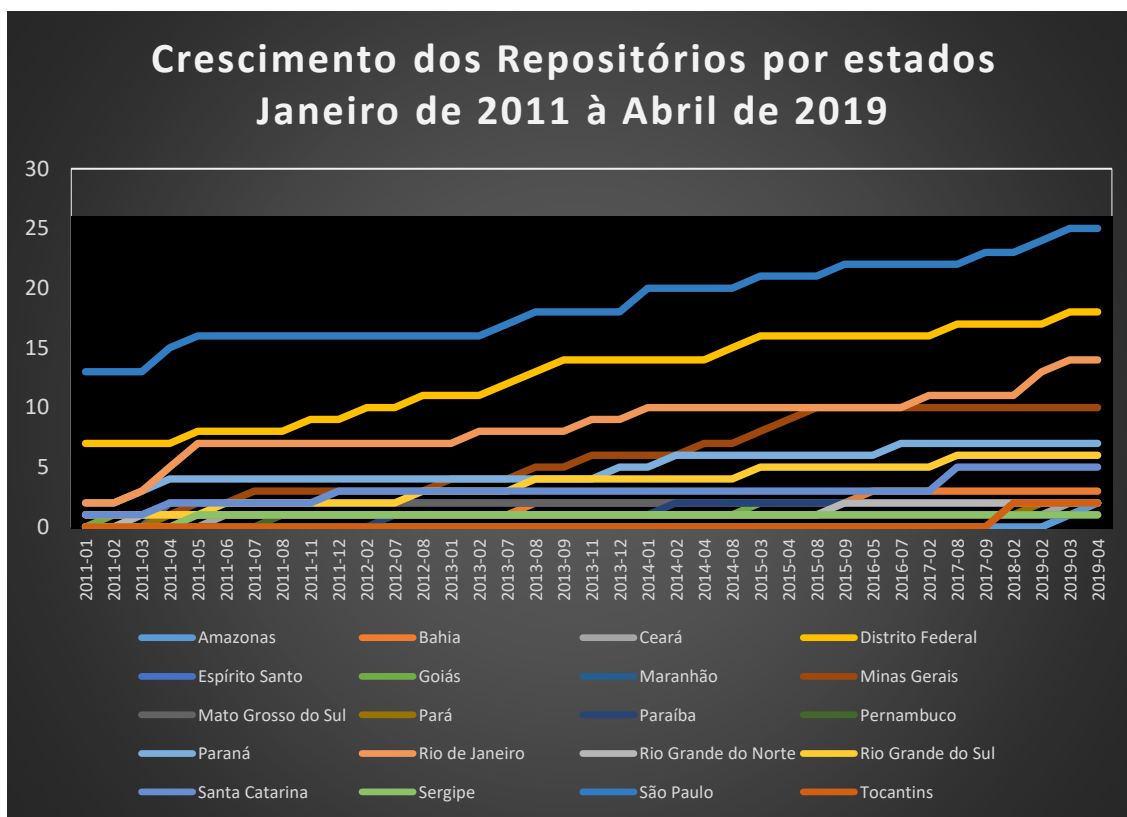


Figura 3: Crescimento de repositórios brasileiros por Unidade Federativa no OpenDOAR

Características

Características mais específicas da infraestrutura dos repositórios brasileiros registrados podem ser identificadas a partir dos dados do OpenDOAR. É possível categorizar tipos de repositório (institucional, disciplinar, etc.), cobertura do assunto (multidisciplinar ou assunto específico), tipo de conteúdo registrado (artigos de periódicos, teses etc.), idioma e tamanho. Esses dados podem ser comparados de forma temporal, para identificar quaisquer alterações importantes desde então.

Em relação a tipologia, a quantidade de repositórios denominados institucionais pelo OpenDOAR é expressivamente maior (Figura 4): 89 repositórios institucionais de um total de 109, correspondendo a 81,6% dos repositórios brasileiros registrados no OpenDOAR. Outras duas importantes categorias de repositórios numeraram bem menos: Repositórios Governamentais foram registrados 7 vezes (6,4%) e Repositórios Disciplinares 11 (10,1%). Somente 2 repositórios (1,8%) foram classificados como Agregadores. É interessante comparar estes percentuais com o registrado no início de 2011. Em janeiro de 2011, Repositórios Institucionais ainda não tinham alcançado ampla maioria, dos 28 repositórios cadastrados, 16 (57,1%) foram categorizados como institucionais.

Repositórios Disciplinares tinham 28,6% (8 repositórios), seguidos por 3 Governamentais (10,7%) e 1 Agregador (3,6%).



Figura 4: Classificação de tipos de Repositórios brasileiros pelo OpenDOAR. 2011-2019

Em se tratando de cobertura de assunto, o registro do mês de abril de 2019 coloca praticamente metade dos repositórios como multidisciplinares, o que não é surpresa, visto o domínio em quantidade dos repositórios institucionais. A Tabela 1 mostra que Repositórios multidisciplinares representam 69 (47%) dos repositórios brasileiros.

Onde os repositórios têm uma cobertura de assunto identificável específica (em vez de ser multidisciplinar), os assuntos são registrados no OpenDOAR. Há uma grande variedade de assuntos listados sem que qualquer área temática principal esteja dominando: todos estão abaixo de 10% do número total de repositórios.

Assunto	Frequência	Porcentagem
Multidisciplinar	69	47%
Saúde e Medicina	8	5%
Direito e Política	8	5%
Ciências Gerais	7	5%
Educação	6	4%
Computação e TI	5	3%
Biblioteconomia e Ciência da Informação	4	3%
Administração e Planejamento	4	3%

Física e Astronomia	4	3%
Tecnologia	3	2%
Negócios e Economia	3	2%
Agricultura, Alimentos e Veterinária	3	2%
Química e Tecnologia Química	3	2%
Artes e Humanidades	2	1%
Geografia e Estudos Regionais	2	1%
História e Arqueologia	2	1%
Ciências Sociais	2	1%
Psicologia	2	1%
Geologia e Ciências Planetárias	2	1%
Arquitetura	1	1%
Engenharia Civil	1	1%
Engenharia Elétrica e Eletrônica	1	1%
Engenharia Mecânica e de Materiais	1	1%
Belas Artes e Artes Cênicas	1	1%
Linguagem e literatura	1	1%
Biologia e Bioquímica	1	1%
Ecologia e Meio Ambiente	1	1%
Matemática e Estatística	1	1%

Tabela 1: Cobertura de assunto em Repositórios Brasileiros

No que se refere a tipologia de materiais em repositórios, a figura 5 nos mostra que os dois tipos de conteúdo mais comum para os repositórios brasileiros listados no OpenDOAR são teses e dissertações e artigos científicos: cerca de 63% dos repositórios incluem estes dois tipos de conteúdo. Outros tipos de conteúdo que se destacam são capítulos de livros (presente em 41% dos repositórios) e anais de conferências (29%). Foram identificados outros tipos, como formatos específicos, incluindo material didático, bibliografias, materiais audiovisuais, dentre outros, em menor número de repositórios. A maioria dos repositórios inclui vários tipos de conteúdo que apoiam claramente o processo de comunicação acadêmica e institucional de várias maneiras. Essas proporções não mudaram desde 2011.

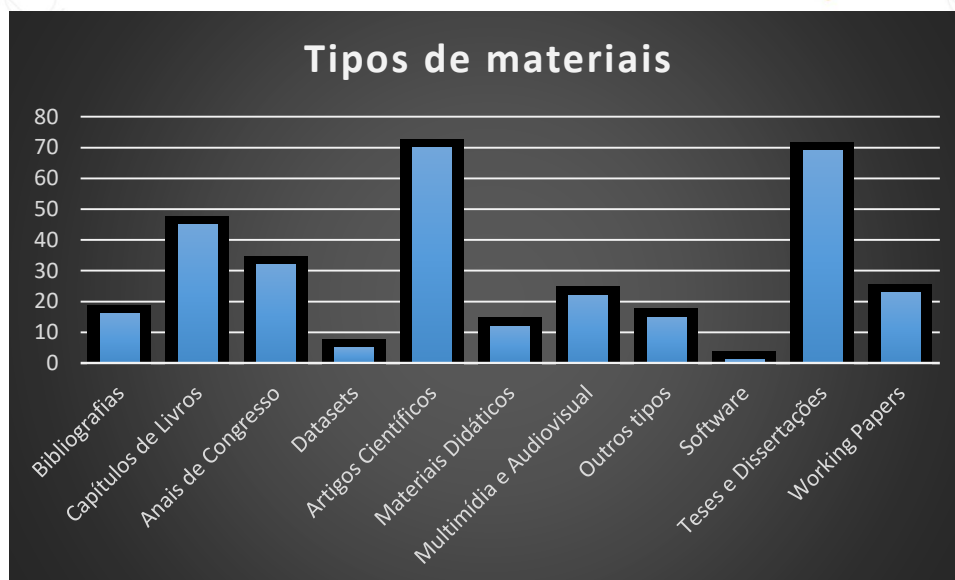


Figura 5: Tipos de materiais hospedados nos Repositórios Digitais Brasileiros

O idioma também é uma característica importante dos repositórios identificados no OpenDOAR. Em abril de 2019, os repositórios brasileiros listados continham recursos em três idiomas. A língua portuguesa é majoritária. 108 repositórios (99%) incluíam material em língua portuguesa (Tabela 2), naturalmente o predomínio do idioma se explica devido a língua nativa do país. O segundo idioma com mais ocorrência é o inglês (32; 29%), que reflete a posição da língua como a globalmente aceita em estudos, particularmente nos domínios da ciência, tecnologia e medicina (Tardy, 2004). Por fim, 16 repositórios (15%) foram listados com ocorrência do idioma espanhol.

Idioma	Total	Porcentagem
Inglês	32	29%
Espanhol	16	15%
Português	108	99%

Tabela 2: Idiomas presentes em Repositórios Brasileiros

Uma questão que merece atenção associada a repositórios é seu tamanho. Grandes repositórios contêm milhões de itens. Por exemplo, o inglês 'Europe PubMed Central' possui, de acordo com o OpenDOAR, 420.000.000 de registros. No entanto a grande maioria dos repositórios é relativamente pequena e isso está de acordo com o contexto brasileiro. O maior repositório brasileiro registrou 275.112 registros. Embora a média de registros em repositórios brasileiros seja 20.451, a Figura 6 nos mostra que a incidência de repositórios entre 100 e 1.000 registros são bem maiores do que em outros grupos. Os dados mostram, em termos simplificados, que há um número muito pequeno de repositórios grandes (ou seja,

com mais de 150.000) e um grande número de pequenos repositórios (com entre 100 e 25.000 itens) que compõem o cenário brasileiro de repositórios.

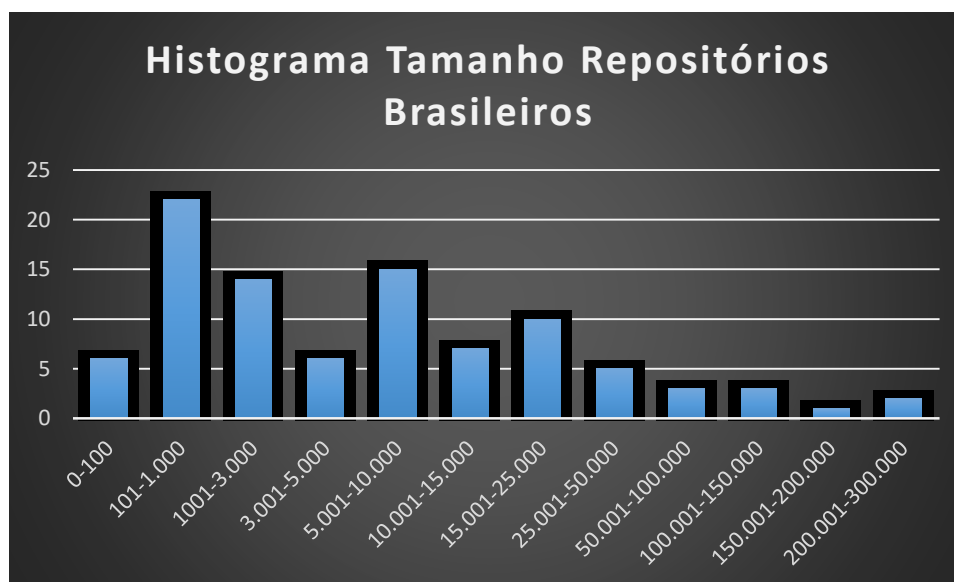


Figura 6: Histograma de frequência de quantidade de itens nos Repositórios Digitais Brasileiros

Ao analisar as plataformas de software utilizadas pelos repositórios, percebemos que o DSpace tem domínio majoritário do mercado e consequentemente permite aos repositórios a utilização de serviços de interoperabilidade com o protocolo OAI. Podemos verificar ainda uma quantidade considerável de soluções desenvolvidas localmente, que dificulta a identificação de seus recursos.

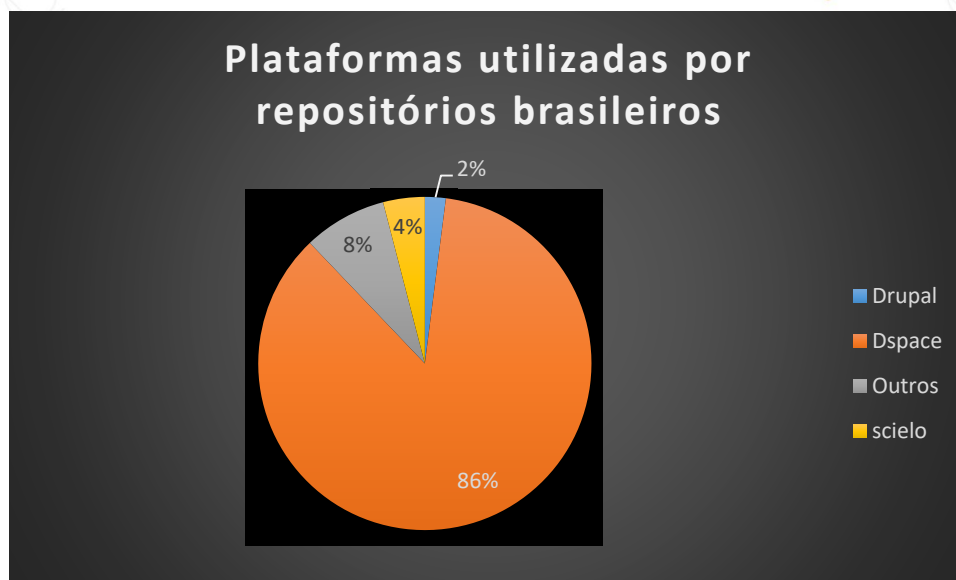


Figura 7: Plataformas de *softwares* utilizados por Repositórios Digitais Brasileiros

Considerações

Os dados apresentados mostram crescimento em repositórios brasileiros entre 2011 e início de 2019. O crescimento foi concentrado na região Sudeste, seguido pelas regiões Centro-Oeste (devido ao forte desempenho do Distrito Federal) e Sul. Somente mais recentemente, houve surgimento de repositórios na região Norte, que possui níveis comparativamente baixos de desenvolvimento de repositórios. Os maiores números de repositórios estão em São Paulo, Distrito Federal e Rio de Janeiro.

Repositórios são predominantemente institucionais em seu foco e multidisciplinar em sua cobertura. Eles ainda têm forte papel como biblioteca de teses e dissertações, porém já têm estabelecido guarda de artigos científicos e, são majoritariamente em português. Embora o tamanho dos repositórios seja difícil de avaliar com precisão, os dados disponíveis indicam que há um pequeno número de grandes repositórios e um grande número de pequenos repositórios. A configuração técnica mais comum para um repositório envolve o uso do software DSPACE (compatível com OAI de código aberto), mas, significativamente, muitas vezes não inclui o licenciamento formal de conteúdo.

Em contexto global, Xia (2012) traz considerações sobre tendências e características da difusão de Repositórios Digitais e Periódicos de acesso aberto. Ele identifica um número de facilitadores e barreiras à adoção que ajudam a explicar muitas das principais tendências apresentadas pelos repositórios brasileiros. Questões de infraestrutura tecnológica podem ser razoavelmente assumidas como barreiras significativas para o desenvolvimento de repositórios na região Norte, por exemplo.

Aparentemente ainda existe um potencial considerável para a continuação da expansão da infraestrutura do repositório no Brasil mesmo que o número de repositórios em determinadas regiões esteja mais próximo de um ponto de saturação. Por exemplo, em São Paulo, todas as grandes instituições de pesquisa agora têm repositórios e, portanto, é provável que o crescimento no número de repositórios futuro não seja o mesmo - embora os dados aqui apresentados ainda não sugiram que isso pode ter começado a acontecer. Isso, é claro, se aplica ao crescimento do número de repositórios em vez de crescimento em seus conteúdos. Como já foi demonstrado, para a grande maioria dos repositórios, permanece um potencial considerável para expandir o conteúdo.

É muito provável que as diferentes regiões continuem a avançar em velocidades diferentes, dependendo do complexo conjunto de fatores técnicos, culturais, gerenciais e econômicos. É importante que tais desenvolvimentos continuem a ser rastreados e monitorados.

Os dados do serviço OpenDOAR apresentados aqui fornecem uma perspectiva importante sobre o desenvolvimento da infraestrutura de repositórios brasileiros na última década. A partir destes resultados, a percepção é de que são necessárias mais pesquisas de análise e comparação com outras regiões do planeta para monitorar tendências de desenvolvimento e avaliar a importância de fatores que os influenciam. O trabalho evidencia a necessidade de fortalecimento de serviços de repositório, particularmente em torno da criação e disponibilização de políticas de reutilização de recursos.

Referências

BAPTISTA, Ana Alice et al. Comunicação científica: o papel da open archives initiative no contexto do acesso livre 10.5007/1518-2924.2007v12nesp1p1. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, p. 1-17, dez. 2007.

BJÖRK, Bo-christer et al. Anatomy of green open access. **Journal Of The Association For Information Science And Technology**, [s.l.], v. 65, n. 2, p.237-250, 6 nov. 2013.

HOUGHTON, John; SWAN, Alma. Planting the Green Seeds for a Golden Harvest: Comments and Clarifications on. **D-lib Magazine**, [s.l.], v. 19, n. 1/2, jan. 2013.

OPENDOAR. **About OpenDOAR**. Disponível em: <http://v2.sherpa.ac.uk/opensoar/information.html>. Acesso em: 26 de março de 2018

PINFIELD, Stephen. Journals and repositories: an evolving relationship?. **Learned Publishing**, [s.l.], v. 22, n. 3, p.165-175, jul. 2009.



TARDY, C. The role of English in scientific communication: lingua franca or Tyrannosaurus rex?. **Journal Of English For Academic Purposes**, [s.l.], v. 3, n. 3, p.247-269, jul. 2004. Elsevier BV. <http://dx.doi.org/10.1016/j.jeap.2003.10.001>.

XIA, Jingfeng. Diffusionism and open access. **Journal Of Documentation**, [s.l.], v. 68, n. 1, p.72-99, 2012. Emerald.

Criação de um repositório temático no contexto das humanidades digitais para gestão da informação

Marcia Teixeira Cavalcanti¹, Ricardo Medeiros Pimenta², Josir Cardoso Gomes³

¹ Universidade Santa Úrsula (USU)/Faculdades Integradas Maria Thereza (FAMATH).
Email: marciacavalcanti@gmail.com

²Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).
Email: ricardo.pimenta@ibict.br

³Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).
Email: josircg@gmail.com

Resumo

Em 2017 foi fundado o Laboratório em Rede de Humanidades Digitais (Larhud), fruto do grupo de pesquisa que se dedica aos estudos críticos e às práticas em Humanidades Digitais. Sua criação teve como objetivo construir um canal estruturado de colaboração entre os pesquisadores internos e externos com foco no desenvolvimento e na reflexão sobre o pensar em soluções metodológicas para a pesquisa nas humanidades. Um caminho adotado pelo Larhud é criar um repositório digital temático, para reunir o conjunto da produção científica e acadêmica dos integrantes do GP, contribuindo para ampliar a visibilidade de sua produção e preservar sua memória. Um repositório pode ser definido como um local de armazenamento, de guarda e arquivamento de objetos digitais. Quanto ao seu propósito, se classifica como institucional – um espaço onde se armazenam os documentos administrativos, acadêmicos e científicos de uma instituição – ou temático – guarda objetos relacionados a uma determinada área sem limites institucionais – cada um com características e estrutura próprias. O artigo propõe apresentar um estudo descritivo, detalhando o passo a passo relacionado à estruturação das etapas de navegação, organização e representação da informação no contexto do repositório pela perspectiva da arquitetura da informação. A etapa inicial foi identificar seus usuários: administradores, pesquisadores e possíveis visitantes. A etapa atual consiste na organização e representação da informação a partir da definição das áreas temáticas, comunidades e coleções que irão compor estruturalmente o repositório, bem como da seleção do tipo de metadados que serão utilizados.

Palavras-chave: Repositório Digital, Humanidades Digitais, Arquitetura da Informação, Preservação Digital.

Abstract

In 2017 was founded the Laboratório em Rede de Humanidades Digitais (Larhud), the result of a research group dedicated to critical studies and practices in Digital Humanities. Its purpose was to build a structured channel of collaboration between internal and external researchers focused on the development and reflection on thinking about methodological solutions for research in the humanities. One way Larhud adopts is to create a thematic digital repository to bring together the scientific and academic production of the members of the GP, contributing to increase the visibility of its production and preserve its memory. A repository can be defined as a location for storing, storing and archiving digital objects. As for its purpose, it is classified as institutional - a space where administrative, academic and scientific documents of an institution - or thematic - store objects related to a certain area without institutional limits - each with its own characteristics and structure. The article proposes to present a descriptive study, detailing the step by step related to the structuring of the steps of navigation, organization and representation of the information in the context of the repository from the information architecture perspective. The initial step was to identify its users: administrators, researchers and potential visitors. The current stage consists of the organization and representation of information from the definition of thematic areas, communities and collections that will structure the repository, as well as the selection of the type of metadata that will be used.

Keywords: Digital Repository, Digital Humanities, Information Architecture, Digital Preservation.

Introdução

Segundo Dacos (2011), as humanidades digitais se referem ao conjunto das ciências humanas e sociais, às artes e às letras e também designam uma transdisciplina, portadora dos métodos, dos dispositivos e das perspectivas heurísticas ligadas ao digital no domínio das ciências humanas e sociais (DACOS, 2011). No Brasil, ainda que a busca pelo tema das Humanidades Digitais se mostre recorrente nos últimos anos, com o discurso da interdisciplinaridade aparecendo em boa parte das iniciativas apresentadas publicamente, é possível percebermos na prática que esta interdisciplinaridade não ocorre de forma efetiva quando o assunto é acesso à informação e reuso de dados gerados por ferramentas e métodos que se mostram

interdisciplinares por natureza.

A virada digital (*digital turn*) abalou, de certa forma, as estruturas do cenário acadêmico, levando professores e pesquisadores a se questionarem e reavaliarem suas estratégias de preservação e de uso da informação científica para a continuidade da produção do conhecimento e da promoção ao seu acesso.

Dentro desse novo cenário que se apresentou foi criado, no ano de 2017, o Laboratório em Rede de Humanidades Digitais (Larhud), fruto do grupo de pesquisa que se dedica aos estudos críticos e às práticas em Humanidades Digitais, mas o cenário também nos convidou a pensar sobre a construção de um repositório que fosse capaz de guardar e preservar dados brutos (*raw data*) provenientes da pesquisa em humanidades de maneira a promover seu acesso pela comunidade científica e que possibilite, entre outras coisas, seu acompanhamento, a troca de informações além de, consequentemente, a geração de inovação e de novos conteúdos. Assim, o Larhud se propôs criar um repositório digital que tem por objetivo reunir o conjunto da produção científica e acadêmica dos integrantes do GP, contribuindo para ampliar a visibilidade de sua produção e também preservar sua memória.

O presente artigo pretende apresentar um estudo de caso descritivo, desenvolvido no âmbito do GP, em que será detalhado o passo a passo relacionado à estruturação das etapas de navegação, organização e representação da informação no contexto do repositório pela perspectiva da Arquitetura da Informação. A etapa inicial foi organizar uma reunião para responder a algumas questões básicas que dariam o suporte para se pensar a estrutura do repositório. O passo seguinte foi identificar seus usuários: administradores, pesquisadores e possíveis visitantes, definindo para cada um deles um nível de permissões. A etapa atual consiste na organização e representação da informação a partir da definição das áreas temáticas, comunidades e coleções que irão compor estruturalmente o repositório, bem como da seleção do tipo de metadados que serão utilizados.

O início de tudo

Um repositório pode ser definido, no contexto atual, como um local de armazenamento, de guarda e arquivamento de objetos digitais. Os repositórios, quanto ao seu propósito, são classificados como institucionais – um espaço onde se armazenam os documentos administrativos, acadêmicos e científicos de uma instituição – ou temáticos – guarda objetos relacionados a uma determinada área sem limites institucionais, como proposto pelo Larhud para a área das Humanidades Digitais – cada um com características e estrutura próprias.

Os repositórios digitais (RDs) são bases de dados online que reúnem de

maneira organizada a produção científica de uma instituição ou área temática. Os RDs armazenam arquivos de diversos formatos. Ainda, resultam em uma série de benefícios tanto para os pesquisadores quanto às instituições ou sociedades científicas, proporcionam maior visibilidade aos resultados de pesquisas e possibilitam a preservação da memória científica de sua instituição. Os RDs podem ser institucionais ou temáticos. Os repositórios institucionais lidam com a produção científica de uma determinada instituição. Os repositórios temáticos com a produção científica de uma determinada área, sem limites institucionais. (IBICT, s.d.)

A ideia de criação de um repositório temático para o Larhud surgiu após os pesquisadores perceberem que estavam gerando um número significativo de dados em suas pesquisas, mas não tinham um local em que estes dados pudessem ser preservados e compartilhados. Mas também ficou perceptível que os dados, enquanto fontes de pesquisa em potencial, não estavam recebendo o tratamento ideal, correndo mesmo o risco de se tornarem irrecuperáveis.

Do "crítico cultural" ao "processador de dados", é fato que este novo contexto de mudança vivida pelo pesquisador das humanidades lato sensu representa um novo "despertar" para um desafio razoavelmente difícil que se baseia no desenvolvimento de competências informacionais aplicadas às novas tecnologias de informação e comunicação (NTIC) por pesquisadores das humanidades. (PIMENTA, 2017, p.12)

Uma das indagações iniciais por parte dos pesquisadores envolvidos na criação do repositório foi sobre a plataforma a ser usada. A dúvida inicial era sobre o uso do Dspace, um software para o gerenciamento de acervo digital, utilizado para implementação de repositórios institucionais, ou o Zenodo, um repositório de dados de pesquisa criado pelo OpenAIRE e pelo *European Organization for Nuclear Research* (CERN) para fornecer um local confiável para os pesquisadores depositarem seus conjuntos de dados. Após algumas reuniões o Zenodo foi escolhido, e a decisão teve como parâmetros norteadores dois elementos principais: (1) disponibilidade de ser um repositório em «nuvem» em uma instituição de pesquisa confiável e não de mercado; (2) ser uma tecnologia de acesso aberto, além de gerar um identificador de objeto digital (DOI) que torna os dados depositados facilmente citáveis.

Posteriormente, considerando o resultado de reuniões ocorridas entre representantes do LARHUD e demais pesquisadores do IBICT Brasília, tomamos conhecimento do posicionamento daquela instituição em utilizar o Invenio – um software de código aberto usado para criação de repositórios, sendo um deles o próprio Zenodo. O apoio do IBICT foi fundamental para decidirmos pelo uso do Invenio na criação de um repositório «desenhado» de acordo com nossas escolhas e que ainda possui as

características do Zenodo descritas acima.

É preciso esclarecer que esta ainda é uma atividade em andamento dentro do Grupo de Pesquisa e, nesse escopo, não somente o «desenho» de como se dará tal repositório como sua implementação estão a ser desenvolvidos com a colaboração de todos os pesquisadores participantes da pesquisa.

Arquitetura da Informação

Vivemos na sociedade da informação, que significa dizer, de forma bem genérica, que vivemos em uma sociedade inaugurada pela internet e caracterizada por constantes mudanças ocasionadas pela tecnologia, o que passa a exigir de nós a habilidade de transformar o mar de informações que recebemos diariamente em conhecimento. Informações estas que nos chegam, ou deveriam chegar, de forma democratizada e universal e que são acessadas por nós via diferentes suportes tecnológicos e espaços informacionais digitais.

Quando pensamos no contexto acadêmico, as informações e os dados que os cientistas buscam encontram-se em sites, plataformas e repositórios, para citar apenas alguns destes espaços. Mas para que estas informações e dados possam ser localizados e recuperados para serem usados, é preciso que estes espaços informacionais tornem essas ações possíveis, ou seja, possuam “[...] uma arquitetura capaz de otimizar ou dificultar a experiência informacional dos sujeitos nestes ambientes” (GUSMÃO, 2018, p. 4).

A Arquitetura da Informação (AI) é a disciplina, ou deveria ser, que nos auxilia para a construção de espaços informacionais pelo qual são praticadas as ações de sistematizar e organizar a informação para que o usuário possa recuperá-la. Segundo Macedo (2005), o termo foi utilizado pela primeira vez pelo arquiteto Richard Saul Wurman em 1976, que o definia como a ciência e a arte de criar instruções para espaços organizados.

Na visão de Wurman a reunião, a organização e a apresentação da informação serviam a propósitos característicos aos das tarefas da Arquitetura. A Arquitetura da Informação seria uma expansão da profissão da Arquitetura, porém aplicada a espaços de informação. E as estruturas de informação influenciariam interações no mundo da mesma forma que as estruturas dos edifícios estimulam ou limitam as interações sociais. (MACEDO, 2005, p. 105)

Gusmão (2018, p. 3), citando Concklin (1987), destaca que problemas como a desorientação (*disorientation*) e o transbordamento cognitivo (*cognitive overhead*) são recorrentes nos diversos tipos de ambientes de informação digital. O autor também cita Oliveira e Vidotti (2013), que sugerem outros problemas de ordem navegacional, de

organização e de recuperação da informação que podem estar associados à questões qualitativas na Arquitetura da Informação desses ambientes. Partindo destas questões, e baseado naquilo que propõe a AI, começamos a pensar como seria organizada a estrutura do repositório para a promoção da recuperação da informação.

Decidir sobre como serão organizados os objetos digitais que compõem o repositório a ser criado para que não apenas eles estabeleçam uma relação entre si, mas também sejam localizados e recuperados, pode parecer uma tarefa fácil a primeira vista, mas não é. É preciso definir as categorias em que esses objetos serão classificados, como eles serão depositados e como serão estabelecidas estas relações, e a AI será essencial para auxiliar na tomada de decisão.

Como o repositório está sendo pensado

A etapa inicial foi nos reunirmos para responder a algumas questões básicas:

- Qual o objetivo do repositório? Reunir, armazenar, organizar, recuperar, preservar, disseminar.
- O quê? Produção científica e intelectual
- De quem? Docente, pesquisador, técnico, aluno
- Acesso? Livre
- Termos na caixa de busca: autor, ano, título, tema, assunto

Partindo destas questões, começamos a pensar como seriam organizados/estruturados os objetos digitais no repositório para que se realizasse a recuperação da informação. Pela familiaridade com o Dspace, ainda que este não tenha sido o escolhido para ser usado, a forma de pensar a organização foi hierárquica:

↳ Comunidades – contém as coleções e subcomunidades: DISCIPLINAS, CURSOS, FERRAMENTAS são algumas das comunidades sugeridas

↳ Coleções – o local onde se agrupam os documentos por tipo

Mas as comunidades e as coleções ainda não foram nomeadas.

A forma como se dará a organização dos objetos digitais para sua adequada recuperação foi pensada, inicialmente, conforme a imagem abaixo. Nela é possível compreender que dados e demais documentos (produção científica acadêmica, bibliográfica, vídeos, entrevistas em áudio, imagens, gráficos, tabelas), uma vez produzidos pelos atores sociais implicados serão todos encaminhados para o mesmo processo de curadoria de maneira a ser classificado como pertencente ao âmbito, ou não, das Humanidades Digitais. Em caso positivo este conjunto de dados, informações e documentos, passará pelo processo de depósito no repositório ao passo que serão duplicados em um servidor local como segurança.

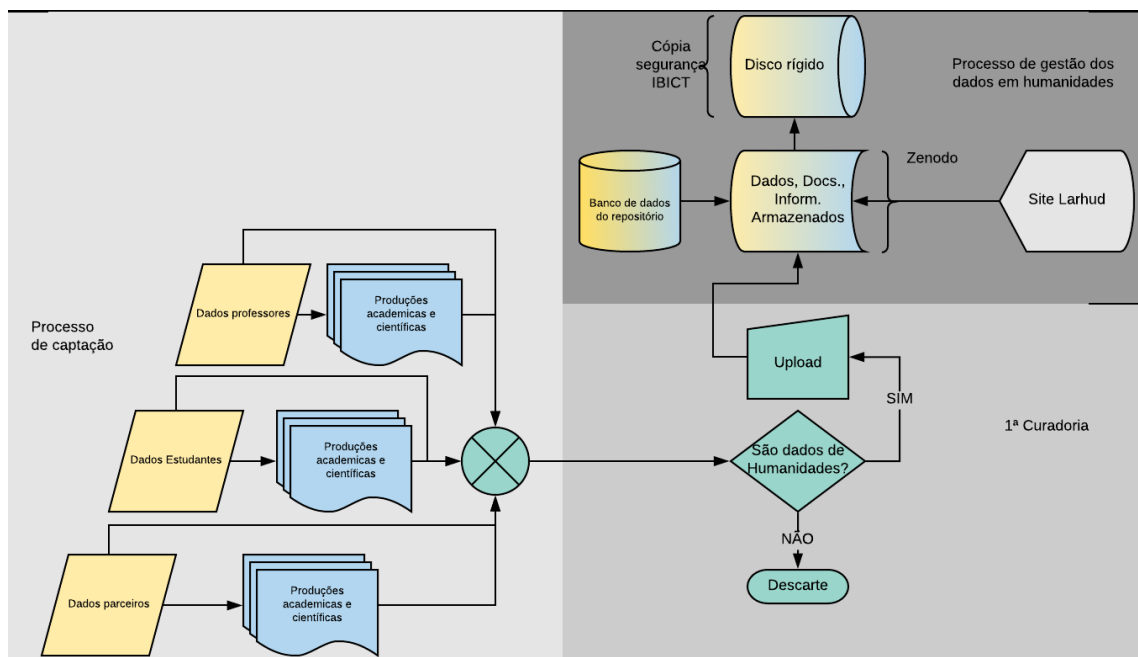
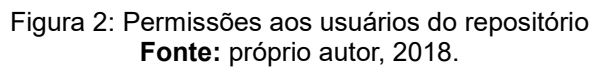


Figura 1: Diagrama de fluxo do repositório Larhud em construção.

Fonte: próprio autor, 2018.

A etapa seguinte foi identificar seus usuários: administradores, pesquisadores e possíveis visitantes; e para cada um deles foi definido um nível de permissões. Inicialmente, aos usuários comuns o acesso ao conteúdo do repositório se dará unicamente pelo sítio eletrônico do Larhud. Por esta entrada o usuário terá acesso ao conteúdo disponível e armazenado/preservado no Zenodo. Os membros do laboratório, no entanto, terão acesso tanto ao Zenodo como ao servidor local. A figura abaixo representa as permissões que serão dadas aos usuários:



Um problema que surgiu durante as discussões do grupo foi como pensar a interface para se acessar o acervo que estará disponível no Zenodo, ou seja, que compõe o repositório em si. Inicialmente essa etapa se dará acessando o site do Larhud, como dito acima. Procuramos diferentes pessoas que conhecem e trabalham com o Zenodo para saber se existe a possibilidade de combiná-lo com outro programa que permita a efetiva organização da informação como descrita até aqui, mas até o momento não conseguimos uma solução para este impasse. Essa questão é de suma importância pois precisamos garantir um acesso fácil, rápido e objetivo para um usuário que, enquanto potencial cientista das ciências humanas, pode não possuir competências digitais informacionais para uma procura profunda e complexa.

Considerações Finais

O momento atual, quando da escrita deste artigo, foi o de idealização, desenho da estrutura e busca por meios materiais de tornar concreto o projeto do repositório. O recurso do repositório em nuvem, apesar de seu servidor residir na União Européia, foi identificado como vantajoso para a implementação desse mesmo repositório destinado às humanidades.

A partir da proposta de criação de um repositório que atenda às Humanidades Digitais, o uso do Zenodo poderá permitir, no escopo das atividades realizadas pelo Larhud, armazenar os resultados de pesquisas científicas multitemáticas e disponibilizar o seu acesso a diferentes usuários, principalmente porque ele permite que a produção seja referenciada via um DOI próprio. Também esperamos que seja possível, a partir desse repositório, desenvolver meios para fortalecer uma comunidade em rede sobre a temática das Humanidades Digitais assim como espessar a produção científica sobre Humanidades Digitais à luz da Ciência da Informação.

Referências

CABEZAS, A. La Referencia and Cern enter into agreement to promote and expand open science in Latin America. LA Referencia [site]. 29 out. 2018. Disponível em: <<http://www.lareferencia.info/es/component/k2/item/218-la-referencia-and-cern-enter-into-agreement-to-promote-and-expand-open-science-in-latin-america>>. Acesso em: 29 set. 2018.

DACOS, M. **Manifesto das Digital Humanities**. That Camp Paris. [blog]. 2011. Disponível em: <<https://tcp.hypotheses.org/497>>. Acesso em: 29 set. 2018.

GUSMÃO, F. C. M. Elementos de arquitetura da informação no Repositório Eletrônico Institucional da UFPB. **Rev. Inf. na Soc. Contemp.**, Natal, RN, Número Especial, 2017. Disponível em: <<https://periodicos.ufrn.br/informacao/article/view/12282>>. Acesso em: 29 set. 2018.

INSTITUTO Brasileiro de Informação em Ciência e Tecnologia – IBICT. **Sobre Repositórios Digitais**. Disponível em: <<http://www.ibict.br/informacao-para-ciencia-tecnologia-e-inovacao%20/repositorios-digitais>>. Acesso em: 28 out. 2018.

MACEDO, F. L. O. **Arquitetura da Informação: aspectos epistemológicos, científicos e práticos**. Brasília: CID/UnB, 2005. 190 fl. (Dissertação de mestrado).



Bibliografia Consultada

YAGUI, M. M. M., VIEIRA, A. C. M., DA CRUZ, S. M. S. **Arquitetura e implementação de um repositório para plantas medicinais em uso no Brasil**. Disponível em: <https://www.researchgate.net/publication/311841590_Arquitetura_e_implementacao_d_e_um_repositorio_para_plantas_medicinais_em_uso_no_Brasil> . Acesso em: 13 ago. 2018.

Repositorio digital de micrografías electrónicas: Una herramienta para la exploración y divulgación de micromundos ocultos

Cecilia Aymará Gallardo¹, Luciano José Martínez², Manuel Siñeriz³, Virginia Helena Albarracín⁴

¹ Técnica Universitaria en Fotografía por la Facultad de Artes de la Universidad Nacional de Tucumán. Abogada por la Facultad de Derecho de la Universidad Nacional de Tucumán. Técnica asociada de la Carrera del Personal de Apoyo a la Investigación del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), con lugar de trabajo en el Centro de Investigaciones y Servicios en Microscopía Electrónica (CISME) CONICET UNT. Coordinadora de ReD-CISME.

Correo: cime.com@tucuman-conicet.gov.ar

² Farmacéutico y Bioquímico por la Facultad de Química, Bioquímica y Farmacia de la Universidad Nacional de Tucumán. Profesional adjunto de la Carrera del Personal de Apoyo a la Investigación de CONICET, con lugar de trabajo en CISME. Responsable técnico CISME. Correo: cime.meb@tucuman-conicet.gov.ar

³ Biólogo y Doctor en Bioquímica por la Facultad de Ciencias Naturales e Instituto Miguel Lillo. Profesional adjunto de la Carrera del Personal de Apoyo a la Investigación de CONICET, con lugar de trabajo en CISME. Correo: cime.met@tucuman-conicet.gov.ar

⁴ Bióloga, Doctora en Bioquímica, Microbióloga Molecular, Investigadora Independiente CONICET, Directora Centro de Investigaciones y Servicios Microscopía Electrónica (UNT-CONICET), JTP Catedra de Biología Molecular Fac. De Ciencias Naturales e IML, UNT. Profesora Titular Catedra de Biología Molecular e Ingeniería Genética, Instituto de Salud y Calidad de Vida, Universidad de San Pablo-T. Correo: cime@tucuman-conicet.gov.ar

Resumen

Introducción: Con el fin de realizar una descripción sistemática se propone digitalizar y curar el acervo de micrografías obtenidas por microscopía electrónica que conserva la institución a fin de constituir un catálogo de datos micrográficos para la educación y divulgación científicas. **Metodología:** Consiste en organizar y realizar la conversión analógica-digital de la colección de acuerdo a lo siguiente: i) desarrollo de protocolos para la digitalización; ii) descripción de la colección conforme estándares internacionales; iii) diseño de sistema de almacenamiento; iv) maquetado, diseño y puesta en línea de una plataforma virtual; v) descripción de las micrografías en dicha plataforma. **Resultados:** El principal producto que genera la institución son micrografías obtenidas con microscopios electrónicos de transmisión y barrido y, reunido el acervo, se contabilizan 40.000 micrografías nativas analógicas y 27.700 digitales, habiéndose digitalizado, al momento, el 30,5% del

mismo. El material a catalogar es totalmente diverso dentro de las ciencias donde un 20 % corresponde a biopsias renales y de piel, 57,6% a material biológico en general y el resto a muestras no biológicas. Se encuentra en desarrollo y diagramación la plataforma web con diferentes sistemas de acceso. **Conclusiones:** La concreción de este repositorio se erige como pionero en el área ya que es el primero especializado en micrografías dentro de la región. Su puesta a disposición en línea permitirá un acceso programado a un recurso valioso para la educación y divulgación científica.

Abstract

Introduction: In order to carry out a systematic description, it is proposed to digitize and cure the collection of micrographs obtained by electronic microscopy that the institution conserves in order to constitute a catalog of micrographic data for scientific education and divulgation. **Methodology:** Consists of organizing and carrying out the analog-digital conversion of the collection according to the following: i) development of protocols for digitization; ii) description of the collection according to international standards; iii) storage system design; iv) layout, design and putting online a virtual platform; v) description of the micrographs in this platform. **Results:** The main product generated by the institution are micrographs obtained with transmission and scanning electron microscopes and, gathering the collection, 40,000 native analog micrographs and 27,700 digital micrographs were recorded, 30.5% of which were digitized at the time. The material to be catalogued is totally diverse within the sciences where 20% corresponds to kidney and skin biopsies, 57.6% to other type of biological material and the rest to non-biological samples. The web platform with different access systems is under construction. **Conclusions:** The concretion of this repository stands as a pioneer in the area since it is the first specialized in micrographs within the region. Its availability online will allow scheduled access to a valuable resource for education and scientific dissemination.

Palabras clave: Repositorio – micrografías – microscopía electrónica

Objetivos

El proyecto de repositorio digital del Centro de Investigaciones y Servicios en Microscopía Electrónica (CISME) se propone organizar, digitalizar y catalogar el acervo de micrografías obtenidas por microscopía electrónica, tanto de transmisión como de barrido para generar un catálogo de datos microscópicos para su uso con fines educativos, de investigación y divulgación científica.

Los objetivos específicos son:

- Reunir, preservar y difundir la producción micrográfica de la institución.

- Aumentar la visibilidad de la producción científica y datos recolectados de las instituciones implicadas.
- Poner a disposición recursos científicos y educativos de alto valor a toda la comunidad interesada.
- Conectar el repositorio digital institucional en bases de datos regionales, nacionales y latinoamericanos.

Fundamentación

El Centro de Investigaciones y Servicios en Microscopía Electrónica (CISME) –dependiente del Consejo Nacional de Investigaciones Científicas y Técnicas y la Universidad Nacional de Tucumán- presta servicios consistentes en la visualización y análisis de muestras a elevadas magnificaciones y altas resoluciones con microscopios electrónicos de transmisión y barrido y sus principales resultados, las imágenes registro de ese análisis, son denominadas micrografías.

La microscopía electrónica permite observar especímenes a altos aumentos, superando ampliamente la capacidad visual del ojo humano y de los microscopios ópticos. Para la formación de la imagen utiliza un haz de electrones, con longitud de onda más pequeña que un haz de fotones, por lo cual puede llegar a grandes resoluciones. En la microscopía electrónica de barrido (MEB), ese haz de electrones, al hacer un recorrido por la superficie de la muestra brinda una señal que permitirá visualizarla con un gran nivel de detalle y profundidad de campo. Las imágenes obtenidas por microscopía electrónica de transmisión, en cambio, muestran la estructura interna de la muestra, es decir, la ultraestructura. Estas tecnologías permiten magnificaciones de entre 1.000.000 a 10.000.000X.

El Centro cuenta con cuatro microscopios electrónicos, de los cuales dos obtienen imágenes de manera analógica: el microscopio electrónico de barrido Jeol 35CF y el microscopio electrónico de transmisión Zeiss EM109. En los últimos años se han incorporado dos microscopios de última generación -el microscopio electrónico de barrido Zeiss SUPRA 55VP y el microscopio electrónico de transmisión Zeiss Libra 120, único en el país-, que han posibilitado un gran salto de calidad tecnológica en cuanto a calidad de imagen, resolución y aumentos alcanzados.

Este archivo de imágenes, que denominamos micrografías por ser las que se obtienen de objetos no visibles a simple vista mediante la ayuda de instrumentos ópticos o electrónicos como lupas y microscopios (Langford, 1977), conforman un patrimonio científico y educativo que requiere de su conservación y custodia. Para ello, se requiere, el diseño de estrategias de gestión de imágenes y difusión que permitan su perdurabilidad en el tiempo, su reutilización por las nuevas generaciones y, en su caso, la generación de valor agregado.

Es reconocida la importancia que revisten los repositorios digitales institucionales en tópicos como unidad, libre acceso de contenidos, preservación, divulgación de la obra, interoperatividad, etc. En Argentina, dicha importancia ha

sido consagrada legislativamente, a través de la sanción de la Ley 26899 del año 2013 que establece que las instituciones del Sistema Nacional de Ciencia y Tecnología, que reciban financiamiento del Estado Nacional, deben crear repositorios digitales institucionales de acceso abierto y gratuito en los que se depositará la producción científico-tecnológica nacional.

En ese contexto, es que nuestro grupo lleva adelante una serie de tareas y procedimientos dirigidos a instrumentar el Repositorio Digital de Micrográficas de nuestro Centro de Investigaciones y Servicios (ReD-CISME). La creación del mismo fue aprobada por decisión del Consejo Directivo del Centro Científico Tecnológico CONICET Tucumán en fecha 23 de febrero de 2016, quedando a cargo de la coordinación del mismo la encargada del área Fotografía y Comunicación del Centro, gestora del proyecto, la Técnica en Fotografía Cecilia Gallardo, bajo la dirección de la responsable científica del centro e investigadora Dra. Virginia Helena Albarracín.

Metodología

Para la conformación del repositorio digital de micrográficas electrónicas se organizó y digitalizó la colección fotomicrográfica del Centro consistente en ca. 60.000 imágenes que constituyen el repositorio que llamamos ReD-CISME obtenidas por el uso de microscopios electrónicos propios: MEB Jeol 35CF y Zeiss SUPRA 55VP y MET Zeiss EM109 y Zeiss Libra 120, siguiendo protocolos estandarizados para la preparación de muestras (Albarracín et al., 2008; Rasuk et al., 2017; Zannier et al.2019).



Fig. 1 MET Zeiss EM109



Fig. 2 MEB Jeol 35CF



Fig. 3 MEB Zeiss SUPRA



Fig. 4 MET Zeiss LIBRA 120

Se utilizaron las siguientes metodologías:

i) Descripción de la colección conforme estándares de descripción archivística. Para ello fue necesario, en primer lugar, reunir la colección que, si bien pertenecía a una misma institución alojada en un mismo espacio físico, por la división tareas, el tipo de microscopía utilizada para la recolección de los datos, etc. se encontraba dispersa. Por otro lado, se hizo un relevo de la cantidad exacta de elementos con los que cuenta el archivo ya que esto determina las estrategias de resguardo a implementar, las posibilidades reales en cuanto a recursos, personal y equipo con el que se cuenta. Para ello, es preciso realizar un inventario de la totalidad del acervo, estableciéndose la cantidad de piezas que se deberán procesar en las diferentes áreas: Digitalización, catalogación, archivo y catálogo online.

Cada elemento digital (nativo y digitalizado con posterioridad a la captura) recibe un número de inventario único y progresivo al que se antepondrá la sigla CISME. Se prevé el posible crecimiento del acervo, asignándose un número de dígitos acorde a ello. Se consideran siete dígitos.

ReD-CISME adhiere a las Directrices del Sistema Nacional de Repositorios Digitales, por lo tanto, para el manejo de metadatos, utiliza las características y semántica del esquema Dublin Core Simple, con los campos básicos recomendados y algunos específicos para el repositorio (LA Referencia, 2018).

Así, los campos establecidos son los siguientes:

- Número de inventario
- Ubicación. Se consigna si el archivo se encuentra en el repositorio general o forma parte de algún catálogo especial como CATMED, por ejemplo.
- Tipo de microscopio. Se señala si corresponde a MET, MEB u otro.
- Título de la imagen. A través de la denominación de la imagen se deberá expresar una idea sintética de lo que se observa en la imagen que será útil al usuario al momento de la búsqueda y recuperación del archivo.
- Resolución
- Autor.
- Descripción
- Licencia.
- Materia. Se establece a qué categoría científica pertenece el objeto digital. Así, hasta el momento, la categorización se ha establecido en las siguientes grandes áreas: Ciencias de la vida, ciencias de la tierra, ciencia de los materiales, nanopartículas/nanotubos y aleaciones. A su vez, cada una contempla subcategorías. Por ejemplo, dentro de ciencias de la vida encontramos virus y procariotas, las cuales a su vez absorben otras subcategorías.
- Fecha de captura, es decir de recolección del dato.
- Tipo. Se establece si el dato pertenece a algún tipo de trabajo en particular: tesis doctoral, trabajo de investigación, artículo, capítulo de libro, libro, etc.
- Editor, en el caso que correspondiera.
- Colaborador.
- Formato.
- Identificador o DOI (Digital Object Identifier)
- Fuente. Se utiliza para relacionar un objeto que es parte de un recurso mayor. Por ejemplo, cuando varias imágenes pertenecen a una misma muestra.
- Leguaje
- Relación. Utilizada para relacionar a distintas versiones de un mismo objeto.

ii) Desarrollo de protocolos de digitalización de calidad de los originales físicos y nativos digitales. Para el primer caso, el Centro cuenta con escáneres planos de negativos de última generación –Epson Perfection V600 photo- que permiten positivar imágenes en el formato de origen -120mm- rescatando escala tonal y contraste de las micrografías blanco y negro características de la microscopía

electrónica. Optamos por digitalizar las micrografías en formato TIFF, a 16 bits, en una resolución de 300 pixeles por pulgada, con un tamaño estándar de 2000 x 3000 pixeles. Igual protocolo se emplea para la digitalización en el segundo caso, es decir, para las micrografías nativas digitales.

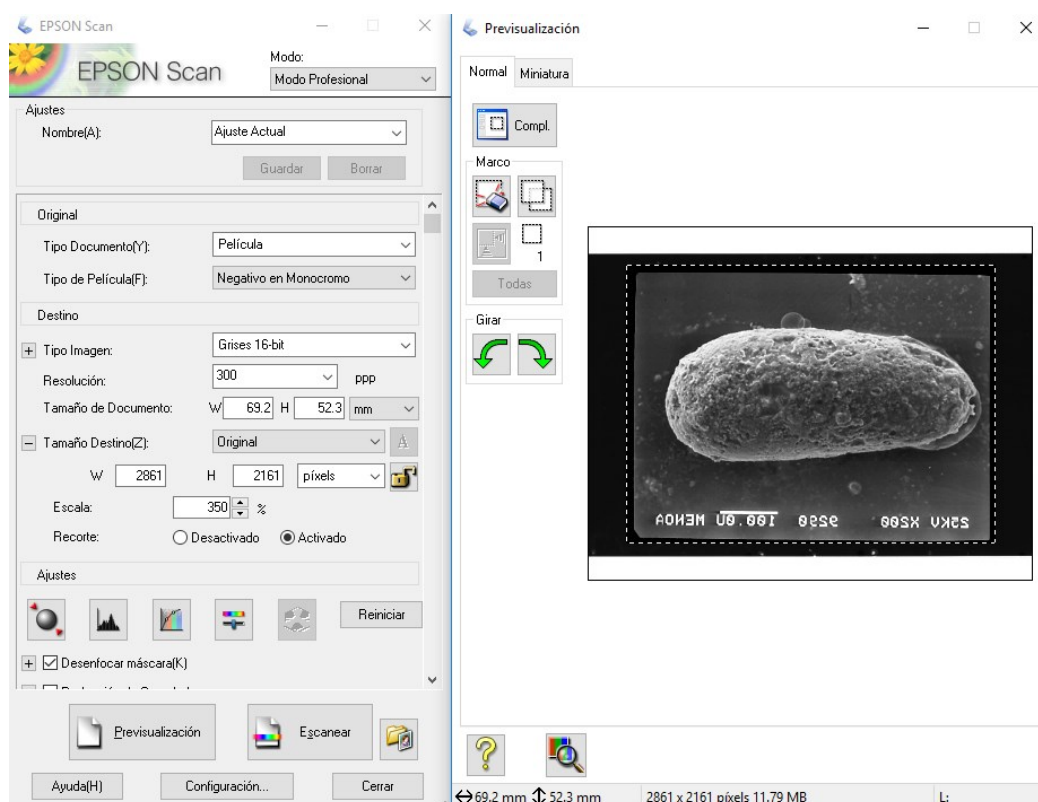


Fig. 5. Vista del software Epson Scan utilizado para el escaneo de películas negativas.

iii) Diseño de sistema de almacenamiento eficaz. Por el volumen de datos que se espera manejar, se ha desarrollado un sistema de almacenamiento que permita, por un lado, repartir las cargas en diferentes equipos, y por otro que facilite la protección de los datos almacenados, evitando sobre todo el almacenamiento en “equipo de trabajo”. Se dispone de doble *back up* en discos rígidos externos, clasificados por materia y fecha, con diferentes localizaciones. Asimismo, se prevé la contratación de un servidor propio con servicio de *back up*.

v) Diseño y puesta en línea de una plataforma web que permitirá el libre acceso de la información según las normas vigentes de propiedad intelectual: Se encuentra en

desarrollo OPEM (*Online Platform for Electron Microscopy Services and Micrographs*) que prevé la puesta en línea del repositorio institucional pero además permitirá ofrecer los servicios de microscopía electrónica del Centro en todo el mundo mediante una plataforma en línea que procesará los pedidos, ayudará a entregar las muestras y proporcionará las micrografías al usuario. Como servicio adicional, la plataforma funcionará como banco de micrografías a demanda para entregar diferentes tipos de productos al usuario solo interesado en la imagen, sin la posibilidad de enviar la muestra.

vi) Generación de catálogos temáticos y descripción de las micrografías con información científica-técnica. En este sentido, se ha desarrollado “CADMED, Catálogo de micrografías electrónicas como recurso para la educación médica”. En este catálogo, se describe un sub-set de microfotografías que pueden ser de interés biomédico (microfotografías de biopsias principalmente). Su puesta a disposición en la plataforma web permitirá un acceso programado a un recurso valioso para la educación médica y capacitación continua, especialmente para anatomo-patólogos. La descripción de toda la colección permitirá ir generando nuevos catálogos acorde a la relevancia y caudal.

Resultados

Al momento se cuenta con 40.000 fotografías en soporte flexible y 27700 nativas digitales.



Fig. 5. Micrografía de cabeza de hormiga obtenida por MEB. Microscopio Zeiss EM109 CISME CONICET UNT. Esta es la primera captura de una imagen por MEB realizada en el Centro en fecha 29 de agosto de 1983

Del total, un 20 % corresponde a biopsias renales y de piel, 57,6% a material biológico en general y el resto a muestras no biológicas. Dentro de las muestras biológicas se describen virus, procariotas: bacterias y arqueas, eucariotas: mamíferos, aves, peces, reptiles, anfibios, Cnidarios/Poríferos, hongos, levaduras, protozoos, plantas, algas, organelas, macromoléculas, cabellos, piezas dentarias, cálculos y alimentos. Las no biológicas se corresponden con elementos arqueológicos, minerales, rocas, suelos, sedimentos, nanopartículas, etc.

Fig. 6. Micrografía de biopsia de piel obtenida por MET Microscopio **Zeiss EM109- CISME CONICET UNT.** Fecha de captura 26 de abril 1994.

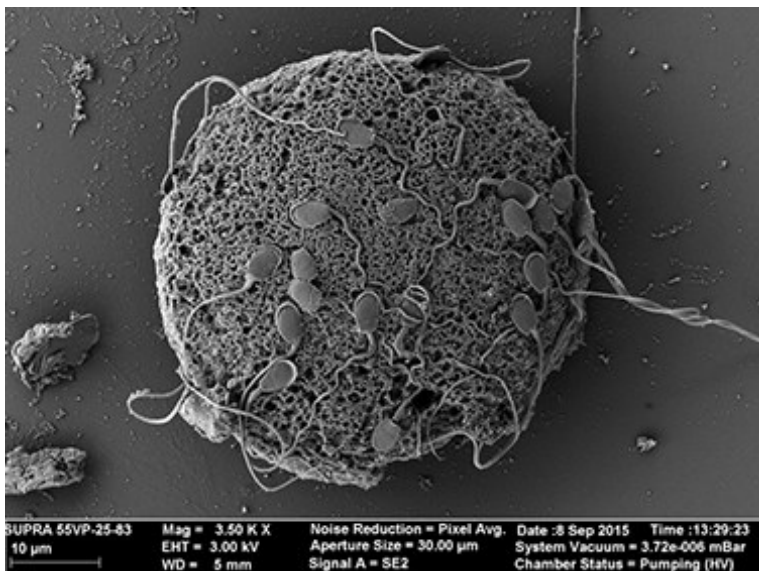
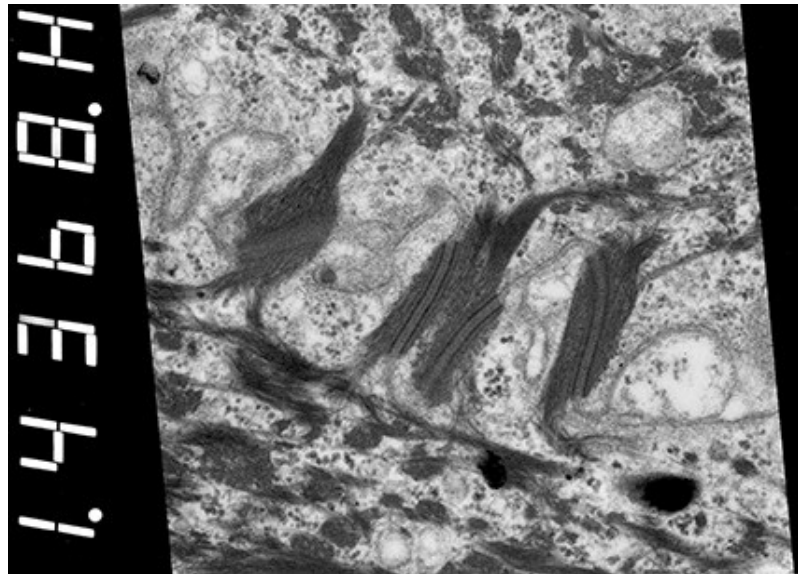


Fig. 7. Imagen de óvulo con espermatozoides de rata obtenida por MEB. Microscopio **ZEISS SUPRA 55VP - CISME CONICET UNT.** Fecha de captura: 8 de septiembre de 2015

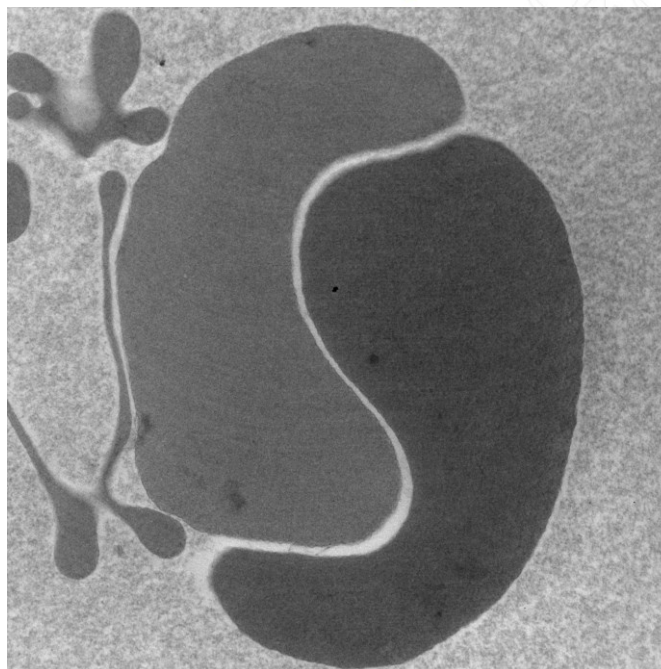


Fig. 8. Micrografía de glóbulos rojos obtenida por MET. Microscopio **Zeiss LIBRA 120 - CISME CONICET UNT.** Fecha captura 17 de septiembre /2018.

Se digitalizaron y catalogaron 12.200 imágenes en alta calidad (30,5 % del total).



Fig. 9. Vista de película negativa montada en carrete de escaneo.

Para su utilización y visualización se diseñó y desarrolló una plataforma *web responsive* con diferentes sistemas de acceso. La misma es capaz de adaptarse para su visualización en cualquier dispositivo. Presenta un panel *web admin* desde donde se podrán importar en lote las imágenes y administrar los servicios del Centro. Todos aquellos que quieran utilizarla deberán registrarse como usuarios. La página web introduce distintas secciones: a) Un banco de imágenes general, donde se encuentran depositadas las micrografías curadas y algunas catalogadas de forma temática por ejemplo CATMED (Catálogo Micrográfico de Interés Biomédico), b) descripción de servicios disponibles. Se encontrará un despegable que permitirá al usuario elegir qué tipo de servicio requerirá tales como entre los ofrecidos por el centro. (Observación en microscopía electrónica de transmisión y de barrido, preparación de muestras, control pericial, procesamiento de imágenes, etc.) c) información y noticias del Centro

Fig. 10. Captura de pantalla de primera entrega de maquetado de página web.

The screenshot displays the CIME Tucumán website interface. At the top, there is a navigation bar with links for 'Banco Imágenes', 'Catálogos', 'Sobre Nosotros', 'Servicios', and 'Contacto'. A search bar and a language selector (English | Español) are also present. The main content area features a form for uploading an image, with fields for 'Título de la Imagen', 'Descripción corta', 'Color' (set to 'Blanco y negro'), 'Tamaño' (set to '2080 X 680 px'), and 'Microscopio' (set to 'TEM'). Below the form are buttons for 'Añadir al carrito' and 'Añadir a Favoritos'. A 'Ver Metadatos' dropdown is visible, which has opened a modal window. This modal contains two columns of metadata fields: the left column includes 'Autor: Cecilia Gallardo', 'Licencia:', 'Materia:', 'Descripción:', 'Editor:', 'Fecha:', 'Colaborador:', and 'Tipo:'; the right column includes 'Formato:', 'Identificador:', 'Tipo:', 'Formato:', 'Identificador:', 'Fuente:', 'Lenguaje:', 'Relación:', and 'N° Inventario:'. At the bottom of the page, there is a footer with a 'MAPA' section containing links to 'Quiénes Somos', 'Nuestro Servicio', 'Banco de Imágenes', and 'Seguimiento', and a 'CONTACTA' section with social media icons and a 'Desarrollado por' credit.

El banco de imágenes cuenta con un buscador y filtros según especificidad tales como rama del conocimiento (ciencias biológicas y de la salud, ciencias de la tierra y minería, ciencia de los materiales, industria farmacéutica, textil, metalmecánica, control pericial y ambiental) y origen de la muestra (virus, bacteria, hongos, etc). Para el depósito de las imágenes en el Repositorio se establece autorización previa del titular de la muestra mediante acuerdo firmado o el transcurso del tiempo legal desde la recolección del dato -5 años-, bajo una licencia no exclusiva, irrevocable y universal, que no incluye derechos de explotación comercial. El usuario del Repositorio Digital podrá descargar las imágenes de manera gratuita, previo *login* como usuario. El registro de los usuarios permitirá realizar análisis estadísticos a posteriori.

El CISME es un centro de investigación y servicios que depende administrativa y financieramente de la Universidad Nacional de Tucumán y del CONICET. Para la concreción del proyecto ReD-CISME se cuenta con subsidios especiales otorgados por: i) el Superior Gobierno de la Provincia de Tucumán, para el diseño, maquetado y puesta en línea de la página web; ii) la Fundación Bunge &

Born y CONICET, para la realización de las actividades de conservación, digitalización, catalogación y la compra de equipamiento y iii) el Consejo Federal de Cultura en el Programa Formación, para llevar a cabo una capacitación en conservación de patrimonio fotográfico y digitalización para el personal involucrado.

Conclusiones

ReD-CISME constituye el primer y más completo repositorio digital de micrografías de la región, siendo un recurso único en su tipo. Su puesta a disposición en línea permitirá, a la comunidad científica, universitaria y al público en general disponer de un valioso recurso para acceder a micromundos ocultos e inadvertidos para el ojo humano.

Referências

Barra Moulain, P. A.; Normas Catalográficas Conaculta; INAH; México; 2005.
Suber P.; Acceso abierto, UAEM, México 2015.

Valdez Marín, J. C.; Conservación de Fotografía histórica y contemporánea. Fundamentos y procedimientos; INAH – CONACULTA; México; 2008.

Valdez Marín, J. C.; Manual ¿Cómo cuidar mis negativos?; INAH – CONACULTA; México; 2000.

Valdez Marín, J. C.; Glosario de términos empleados en conservación fotográfica; INAH – CONACULTA; México; 2001.

Ward, P. La conservación del patrimonio: carrera contra reloj; The Getty Conservation Institute; 2010.

Csillag Pimstein I.; Conservación de Fotografía Patrimonial; Chile; 2000.

Casellas i Serra, L. E.; La gestión archivística de los fondos y colecciones fotográficas, Las Palmas; 2005.

Langford, M.; Manual de técnica fotográfica. Madrid: Hermann Blume ediciones. pp. 335, 330, 1977.

Directrices de LA Referencia, 05 de abril de 2018.

Gallardo, C.A., Martínez LJ, Mazzuco Cánepa M.P.M, Albarracín, V.H. Repositorio del Centro Integral de Microscopía Electrónica. 20° Reunion de Investigadores en Ciencias de la Salud, 2018.



Albarracín, V. H., Winik, B., Kothe, E., Amoroso, M. J., & Abate, C. M. (2008). Copper bioaccumulation by the actinobacterium *Amycolatopsis* sp. AB0. *Journal of basic microbiology*, 48(5), 323-330.

Rasuk, M. C., Ferrer, G. M., Kurth, D., Portero, L. R., Farías, M. E., & Albarracín, V. H. (2017). UV-Resistant Actinobacteria from High-Altitude Andean Lakes: Isolation, Characterization and Antagonistic Activities. *Photochemistry and photobiology*, 93(3), 865-880.

Zannier, F., Portero, L. R., Ordoñez, O. F., Martinez, L. J., Farías, M. E., & Albarracin, V. H. (2019). Polyextremophilic Bacteria from High Altitude Andean Lakes: Arsenic Resistance Profiles and Biofilm Production. *BioMed research international*, 2019.

La vía verde del acceso abierto en Argentina: evolución y desarrollo de los repositorios en las universidades nacionales¹

Carolina Unzurrungaza¹, Marcela Fushimi², Mónica G. Pené³, Ana María Sanllorenti⁴

¹ Universidad Nacional de La Plata. Facultad de Humanidades y Ciencias de la Educación - Instituto de Investigaciones en Humanidades y Ciencias Sociales (CONICET-UNLP). Correo: cunzu@fahce.unlp.edu.ar

² Universidad Nacional de La Plata. Facultad de Humanidades y Ciencias de la Educación - Instituto de Investigaciones en Humanidades y Ciencias Sociales (CONICET-UNLP). Correo: mfushimi@fahce.unlp.edu.ar

³ Universidad Nacional de La Plata. Facultad de Humanidades y Ciencias de la Educación - Instituto de Investigaciones en Humanidades y Ciencias Sociales (CONICET-UNLP). Correo: mpene@fahce.unlp.edu.ar

⁴ Biblioteca Central Dr. Federico Leloir, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires y Universidad Nacional de Mar del Plata. Facultad de Humanidades. Correo: ana@bl.fcen.uba.ar

Resumen

Desde 2013, en Argentina, está vigente la ley 26.899 que establece el mandato de acceso abierto (AA) para la literatura científica y datos primarios resultado de investigaciones financiadas con fondos públicos, a través de repositorios digitales gestionados y sostenidos por las instituciones del sistema científico y tecnológico nacional. Este trabajo tiene por objetivos conocer la evolución y el estado actual de desarrollo de los repositorios que fueron creados en el ámbito de las universidades nacionales argentinas de gestión pública (UUNN), identificar los grupos involucrados en su gestión y analizar las actitudes de los investigadores argentinos respecto a los repositorios institucionales existentes. Para obtener datos descriptivos de la situación actual de los repositorios, se realizó entre febrero-mayo de 2019 un relevamiento de iniciativas de AA observando en forma directa los sitios web principales de las UUNN y de sus facultades, bibliotecas y áreas potencialmente vinculadas a iniciativas de publicaciones en AA, y se registraron los datos en una matriz ad hoc. Para obtener un panorama de las opiniones, intereses y prácticas de los investigadores, se llevó a cabo una revisión bibliográfica y un análisis de la literatura disponible. Entre las conclusiones se destacan: el impacto de la ley nacional en relación con el crecimiento del número de repositorios y su

¹ Esta ponencia presenta avances de los resultados del proyecto de investigación "[Repositorios digitales de acceso abierto en las universidades nacionales argentinas: Desarrollo, evolución y perspectivas de futuro](#)" (PPID040, Periodo 2018-2019) financiado por la Universidad Nacional de La Plata (Argentina).

adhesión al SNRD, el liderazgo de las bibliotecas universitarias en su desarrollo y mantenimiento, la escasa existencia de repositorios de datos primarios, y el surgimiento de otro tipo de iniciativas no alineadas con los requisitos técnicos del AA que demuestran el interés de las comunidades por difundir en forma amplia su producción académica y científica.

Palabras clave: Repositorios institucionales - Argentina - Universidades Nacionales - Ley de acceso abierto.

Abstract

In Argentina, Law 26.899 has been in force since 2013, establishing the Open Access (OA) mandate for scientific literature and primary data resulting from research financed with public funds, through digital repositories managed and supported by the institutions of the national scientific and technological system. The objectives of this work are to know the evolution and the current state of development of the repositories that were created in the context of the Argentine national universities of public management (NNUU), to identify the groups involved in their management and to analyze the attitudes of Argentine researchers regarding the existing institutional repositories. In order to obtain descriptive data on the current situation of the repositories, a survey of OA initiatives was conducted between February and May of 2019, directly observing the main websites of the NNUU and its faculties, libraries and areas potentially linked to OA publication initiatives, and the data were recorded in an ad hoc spreadsheet. In order to obtain an outlook on researcher's opinions, interests and practices, a bibliographic review and analysis of available literature was undertaken. Among the conclusions, the following stand out: the impact of the national law in relation to the growth of the number of repositories and their adherence to the SNRD, the leadership of university libraries in their development and maintenance, the limited existence of primary data repositories, and the emergence of other types of initiatives not aligned with the technical requirements of the OA but which demonstrate the interest of the communities in widely disseminating their academic and scientific production.

Keywords: Institutional repositories - Argentina - Public Universities - Open Access Law

Introducción

Argentina cuenta con una política pública que establece la denominada “vía verde” como mecanismo para alcanzar el acceso abierto a la producción científica generada con fondos públicos, expresada en la Ley Nacional 26.899 que fue

aprobada en 2013 y reglamentada a fines de 2016. Esta ley determina que la literatura científica y los datos primarios de las investigaciones financiadas con fondos públicos deben estar en acceso abierto (en adelante AA) a través de repositorios digitales que las instituciones del sistema científico y tecnológico nacional estarán obligadas a crear y gestionar. Asimismo, exige a los investigadores el depósito de sus producciones y de los datos primarios que generan en estos repositorios.

En América Latina, además de Argentina, tienen legislaciones similares Perú y México, donde establecen la creación de redes de repositorios institucionales para el depósito de su producción científica. En otros países se ha impulsado también la creación de repositorios y el mandato para hacer pública la producción científica (ej. España) pero no como algo obligatorio. Los mandatos más exigentes en realidad son los de las agencias de financiación de la investigación, aunque no todas obligan al depósito en repositorios -como el caso de los Institutos de Salud de Estados Unidos de América (NIH-USA) en PubMed Central-, sino más bien a la publicación en abierto. Recientemente Larivière y Sugimoto (2018), tras estudiar el grado de cumplimiento de los mandatos de AA por parte de los investigadores de 12 agencias de financiamiento a nivel internacional, encontraron que los mismos eran disímiles, según la agencia evaluada. Entre las conclusiones del trabajo se afirma que más que las diferencias culturales o normativas entre disciplinas, son las agencias de financiamiento y sus políticas de seguimiento y monitoreo las que producen mejores nacionales argentinas de gestión pública² (en adelante UUNN) e identificamos a los tasas de AA para las investigaciones y rescatan que es necesario superar las barreras como la infraestructura y la financiación. Si hay políticas claras y se implementan la estructura y los incentivos adecuados, los investigadores cumplen.

América Latina es una región donde la actividad científica y su comunicación están financiadas, en buena parte, con fondos estatales, lo que debería favorecer el AA de los resultados de investigación que generan sus investigadores, hecho que se ve reflejado en el desarrollo de las revistas que se han anclado en la tradición de las universidades públicas (Babini, 2019). A la vez, en el escenario mundial se vislumbra un cambio radical en el escenario internacional del AA (Piwowar et al., 2018) y la comunicación científica que son resultado, en parte, de la aparición de actores que desafían los modelos tradicionales de publicación, es decir, plataformas para difusión, intercambio y obtención de documentos como las redes sociales académicas del estilo de *ResearchGate* o *Academia*, y los sitios como *Sci-Hub* o *LibGen*.

² Cabe aclarar que -debido a la amplitud y complejidad del sistema de ciencia y tecnología argentino-, esta investigación se enfocó únicamente en las universidades de gestión pública, dejando fuera a las universidades de gestión privada, y también a los organismos específicos que realizan actividades de investigación y desarrollo.

En este contexto, y a cinco años de sancionada la ley argentina de AA, resulta de interés indagar el estado actual y las perspectivas a futuro de los repositorios institucionales en Argentina, considerándolos -desde el enfoque que plantea Guédon (2009)- como artefactos sociotécnicos complejos, lo que permite abordar el objeto de estudio con una mirada interdisciplinaria.

En esta ponencia presentamos los resultados parciales de una investigación en curso en la que se busca contribuir al conocimiento de la vía verde en el sur global, destacando sus características distintivas, logros, dificultades y los desafíos que enfrentan para su integración dentro del circuito de la comunicación académica. Avanzamos, en esta ocasión, en conocer la evolución y el estado actual de los repositorios digitales que fueron creados en el ámbito de las universidades grupos involucrados en su gestión. Asimismo, nos adentramos en el análisis de las actitudes de los investigadores argentinos respecto a los repositorios institucionales (en adelante RIs), como espacio para la difusión de su producción, que son evidenciados en los estudios detectados en la literatura a fin de comenzar a tener un panorama más concreto de sus destinatarios.

Estudios previos sobre la vía verde en Argentina

La situación de los repositorios argentinos ha sido estudiada en varias oportunidades. Como primer trabajo debe mencionarse el texto de De Volder (2008), quien describió y caracterizó 15 repositorios de AA existentes hasta ese entonces, señalando que se trataban de esfuerzos aislados que, en su mayoría, eran liderados por bibliotecas universitarias. Otro trabajo es el de Fushimi y Banzato (2010), quienes realizaron un relevamiento de sitios web de universidades nacionales argentinas, identificando 8 repositorios en funcionamiento y 5 en proyección; asimismo, plantearon que las universidades estaban aprovechando las oportunidades que ofrecía la web para visibilizar y difundir la producción de sus investigadores, aunque las iniciativas no necesariamente respondían a los preceptos del AA. Cabe señalar que tanto De Volder (2008) como Fushimi y Banzato (2010) coincidieron en afirmar que la ausencia de políticas públicas hacía que las iniciativas existentes hasta el momento fueran aisladas, y desarrolladas principalmente por las bibliotecas, con escaso apoyo institucional.

En 2011, el Ministerio de Ciencia, Tecnología e Innovación Productiva de Argentina dio a conocer un informe donde planteaba que el número de repositorios digitales en funcionamiento se había elevado a 23 (más 7 que estaban realizando pruebas y 28 que se declararon en proyección), de los cuales el 78% estaban a cargo de bibliotecas (Argentina. MINCyT, 2011). En ese mismo año, Bongiovani y Nakano (2011) expusieron en su trabajo el caso de la creación de un Sistema Nacional de Repositorios Digitales (SNRD)³, en el marco de las acciones del

³ <http://repositoriosdigitales.mincyt.gob.ar/vufind/> (Consultado el 20/5/2019)

Ministerio de Ciencia, Tecnología e Innovación Productiva de Argentina (MINCyT), y las estrategias desarrolladas en pos de lograr el AA a la producción científica y tecnológica de Argentina, con la idea de conformar una red interoperable de repositorios digitales en ciencia y tecnología a partir del establecimiento de políticas, estándares y protocolos comunes a todos sus integrantes. Actualmente el SNRD cosecha y brinda acceso unificado a más de 207.000 registros de la producción depositada en los 33 repositorios adheridos⁴. La actual Secretaría de Ciencia, Tecnología e Innovación Productiva -Ex Ministerio del mismo nombre⁵- es el organismo encargado de monitorear el SNRD, para lo cual fomenta y financia la creación y desarrollo de repositorios, brinda asesoramiento y controla su avance. Otra investigación posterior, realizada por Pené, Unzuñunzaga y Borrell (2016) tras el análisis de 20 iniciativas disponibles en las UUNN de Argentina, dejó entrever una apropiación local del concepto de repositorio, en muchos casos por parte de las bibliotecas de estas instituciones, para difundir objetos digitales de la más variada tipología -y no necesariamente producción científica-, con la finalidad de reunir, registrar, preservar y difundir el patrimonio y la memoria de la institución y de su comunidad académica.

El estudio más reciente de los repositorios del sistema universitario público argentino es el Fushimi (2018), en el que se mostró que hasta 2015 sólo 20 de las 58 universidades nacionales distribuidas en 7 regiones en todo el país contaba con al menos un repositorio en funcionamiento, la mitad de los cuales estaban concentrados en las regiones metropolitana y bonaerense. Señaló, además, que existían otras 12 universidades con 19 iniciativas y proyectos con un bajo nivel de desarrollo. En total, el estudio identificó 51 repositorios, de los cuales 46 eran institucionales y el resto se repartía entre repositorios temáticos, de fuentes para la investigación y de objetos de aprendizaje. Asimismo, de esos 51, 32 estaban en funcionamiento, 18 en proceso de construcción y 1 se encontraba detenido, mientras que 30 eran generales o centrales y 21 especializados.

En cuanto a los actores involucrados en la construcción y el mantenimiento de repositorios, el trabajo de Fushimi (2018) expuso que las bibliotecas fueron uno de los protagonistas fundamentales liderando el 71% de los registrados hasta 2015, en tanto que el Estado Nacional -mediante el diseño y la implementación de la ley 26.899 y el SNRD- puede considerarse el segundo actor relevante para el crecimiento de la vía verde debido a que se constató un crecimiento en el período 2014-2015 superior al registrado en los siete años anteriores. Asimismo, señaló que los niveles de conducción de las universidades nacionales -en tanto espacios donde se realizan las investigaciones- desempeñaron un rol más bien pasivo, con algunas

⁴ Datos actualizados al 22-4-2019, tomados de <http://repositoriosdigitales.mincyt.gob.ar/vufind/>.

⁵ El Ministerio de Ciencia, Tecnología e Innovación Productiva, creado en 2007, fue degradado a Secretaría de Gobierno en septiembre de 2018 ([Decreto 801/2018](#)), pasando a depender del Ministerio de Educación, Cultura, Ciencia y Tecnología, como parte de un proceso de recorte y achicamiento del Estado.

excepciones, aunque por lo general “dejaron hacer”, acompañando tíbiamente las políticas públicas y el impulso de los bibliotecarios. Las editoriales universitarias, que también son actores afectados por estas políticas y comparten algunos de los objetivos e ideales del AA, en general no participaron en la construcción de los repositorios y, aunque se mostraron atentos a su crecimiento por el impacto que podría causarles, se mantuvieron al margen. Se observaron además diferencias entre los editores de revistas científicas y los que se dedican a la publicación de libros. Los actores señalados como más reticentes fueron los investigadores -en tanto autores de la producción científica- ya que se vieron ajenos al proceso de elaboración de la ley y al desarrollo de los repositorios (Fushimi, 2018).

Metodología

Para obtener los datos que permitieran describir la situación actual de los repositorios, se realizó un relevamiento en busca de iniciativas de AA entre los meses de febrero y mayo de 2019. Se dio inicio a la revisión haciendo una observación directa de los sitios web principales de las UUNN y navegando particularmente: facultades y dependencias, secretaría de investigación, secretaría académica, bibliotecas y áreas potencialmente vinculadas a iniciativas de publicaciones en AA. La información de cada iniciativa encontrada fue registrada en una matriz de datos que se construyó *ad hoc*, acompañada de una guía observacional que posibilitó el trabajo colaborativo entre todas las personas que integran el proyecto de investigación. Las variables relevadas fueron: región, universidad, nombre de la iniciativa, tipo y cantidad de iniciativas halladas, estado, antigüedad y evolución, normativa relacionada, nivel de dependencia, software de gestión utilizado, tipos, formatos y cantidades de documentos incluidos, temáticas comprendidas, así como otras características generales de las iniciativas.

En lo que respecta a la medición del grado de desarrollo de un RI, dado que se trata de un indicador de compleja definición que requiere la consideración de numerosos aspectos, se siguió la propuesta de Fushimi (2018) y se avanzó en conocer cinco aspectos: *tamaño*, medido en cantidad de objetos digitales depositados; *visibilidad*, según su registro en las fuentes reconocidas del sector; *políticas*, existencia de políticas y/o normativas de funcionamiento aprobadas por la institución; *antigüedad*, en base a la fecha de puesta en funcionamiento; y *adhesión al SNRD*, es decir, que hayan tramitado su adhesión al sistema nacional.

Para identificar y analizar los grupos involucrados en la gestión de los repositorios, retomando la teoría sociotécnica, se consideraron los grupos de actores que plantea Guédon (2009, p. 584-585), adaptándolos al contexto local y a las características propias de las universidades nacionales.

Finalmente, para obtener un primer panorama de las opiniones, los intereses y las prácticas de los investigadores argentinos respecto a los repositorios, se llevó a cabo una revisión bibliográfica y un análisis de la literatura disponible.

Categorías definidas

Los estudios previos mencionados en el apartado 2 muestran que las UUNN argentinas adoptan formas muy diversas a la hora de difundir y brindar acceso a su producción académica y científica -más allá de la generación de repositorios y revistas AA, que siempre se consideraron las vías principales del acceso abierto-, motivo por el cual fue necesario definir categorías precisas para su organización (ver tabla 1). En este relevamiento se consideraron *iniciativas de acceso abierto a todas aquellas páginas o portales web que reúnen producción científica y/o académica y la disponen para lectura y/o descarga de forma gratuita y abierta, aunque no adopten una licencia de distribución explícita.*

Nombre	Descripción
Repositorio institucional	Archivo digital abierto gestionado por una institución, en el cual se almacena y disponibiliza en la web principalmente literatura científica y académica generada por su comunidad.
Repositorio temático/ disciplinar	Archivo digital abierto gestionado por una comunidad disciplinar o por diferentes instituciones, en el cual se almacena y disponibiliza en la web principalmente literatura científica y académica en base a un interés temático común.
Repositorio de fuentes de investigación	Archivo digital abierto en el cual se almacenan y disponibilizan en la web documentos que fueron utilizados y/o generados como fuentes por una comunidad con fines investigativos.
Repositorio de datos	Archivo digital abierto en el cual se almacena y disponibiliza en la web conjuntos de datos generados durante los procesos de investigación.
Repositorios de objetos de aprendizaje	Archivo digital abierto en el cual se almacena y disponibiliza en la web principalmente recursos educativos generados por una comunidad con fines de enseñanza-aprendizaje.
Editorial universitaria	Sitio web de la editorial de la institución que incluye contenidos de descarga gratuita y abierta.
Portal de libros abiertos	Sitio web que ofrece acceso a libros editados y/o coeditados por la institución y/o su sello editorial propio con alguna licencia de uso abierta.
Portal de revistas abiertas	Sitio web que ofrece acceso unificado a las revistas editadas y/o coeditadas por la institución y/o su sello editorial propio con alguna licencia de uso abierta.
Iniciativa de acceso abierto no clasificada	Sitio web que reúne y ofrece materiales científicos y/o académicos de la institución para lectura y/o descarga gratuita y abierta, pero no responde a ninguna de las anteriores categorías.

Tabla 1: Categorías usadas para clasificar las iniciativas

En tanto para la variable *nivel de dependencia o pertenencia* de las iniciativas encontradas, se definieron dos categorías: *general*, para indicar que pertenece a toda la Universidad, y, *específica*, para aquellas que responden a facultades, departamentos o áreas en particular.

Asimismo, se utilizó la siguiente categorización para tipificar el *estado* de las iniciativas: *en funcionamiento* para las que se encontraban vigentes, en línea y enlazadas desde un sitio web institucional; *en desarrollo* cuando se encontraban mencionadas y/o anunciadas desde un sitio web institucional, o bien cuando contaban con adhesión provisoria al SNRD pero aún no eran visibles; y *detenidas* cuando la iniciativa estaba vigente y en línea, pero no registraba actualizaciones desde 2017, o el sitio permanecía caído luego de sucesivas visitas.

Fuentes consultadas

Para determinar la *visibilidad* de las iniciativas por fuera de la propia institución, se corroboró la presencia de éstas en diferentes directorios nacionales e internacionales vinculados a recursos de AA (ver tabla 2).

Sigla	Nombre	URL	Observaciones
ROAR	Registry of Open Access Repositories	http://roar.eprints.org/	Es un directorio mundial donde los repositorios pueden registrar sus datos de forma voluntaria.
OpenDOAR	Directory of Open Access Repositories	http://v2.sherpa.ac.uk	Similar al anterior, solo que realiza un chequeo de verificación de autenticidad de los datos registrados por parte del directorio.
BECYT	Recursos de Acceso Abierto Argentino	http://www.biblioteca.mincyt.gob.ar/recursos/accesoabierto?geo=AR	Lista de recursos AA organizada y mantenida por la Biblioteca Electrónica de Ciencia y Tecnología.
SNRD	Sistema Nacional de Repositorios Digitales en Ciencia y Tecnología	Consulta de solicitudes de adhesión al SNRD y a la lista informada por Nakano y Azrilevich (2017, p.9)	El trámite de adhesión es solicitado voluntariamente por las instituciones, y su otorgamiento implica cumplir una serie de requisitos establecidos por el SNRD.
REUN	Red de Editoriales Universitarias	http://www.librouniversitario.com.ar/publisher/	También se consideró la mención de la pertenencia a la REUN en los sitios editoriales.
SND	Sistemas nacionales de datos (biológicos, del mar, etc.)	http://sistemasnacionales.mincyt.gob.ar/	Aplicable sólo a repositorios de datos.
DOAB	Directory of Open Access Books	https://www.doabooks.org/	Aplicable a emprendimientos editoriales.
BDU2	SIU-BDU2 Repositorios digitales	http://cosechador.siu.edu.ar/bdu3/	Recolector de repositorios institucionales argentinos desarrollado y mantenido por el SIU ⁶ .

Tabla 2: Fuentes consultadas para determinar la visibilidad de las iniciativas.

⁶ SIU: Sistema de Información Universitario de Argentina (<https://www.siu.edu.ar/>)

Resultados del relevamiento

Caracterización de las iniciativas encontradas

Se detectaron un total de 152 iniciativas AA en las 65 UUNN argentinas existentes, las cuales fueron clasificadas según las categorías antes explicadas (ver gráfico 1). Más de la mitad de éstas fueron identificadas como repositorios institucionales (34,9%) y portales de revistas (20,4%), vías tradicionales del AA. Le sigue en tercer lugar un grupo de iniciativas no clasificadas que disponen de forma gratuita y abierta en la web una producción científica y académica variada (19,1%), y en cuarto lugar un grupo de editoriales universitarias que difunde, también de forma gratuita y abierta en la web, producciones académicas propias (17,8%).

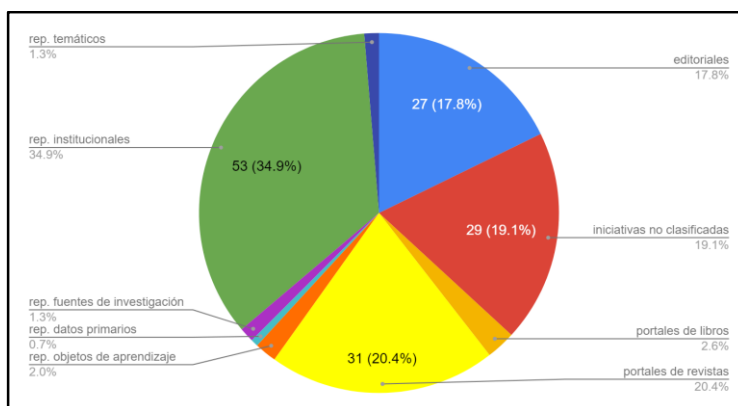


Gráfico 1: Tipos de iniciativas encontradas en las UUNN argentinas en 2019.

Como puede observarse en el gráfico 2, las iniciativas se desarrollan dispersamente tanto a nivel general de la universidad como en niveles inferiores de jerarquía institucional (facultad, centro de investigación o departamento docente). En el caso particular de los 53 RIs hallados, se detectaron 32 generales y 21 específicos.

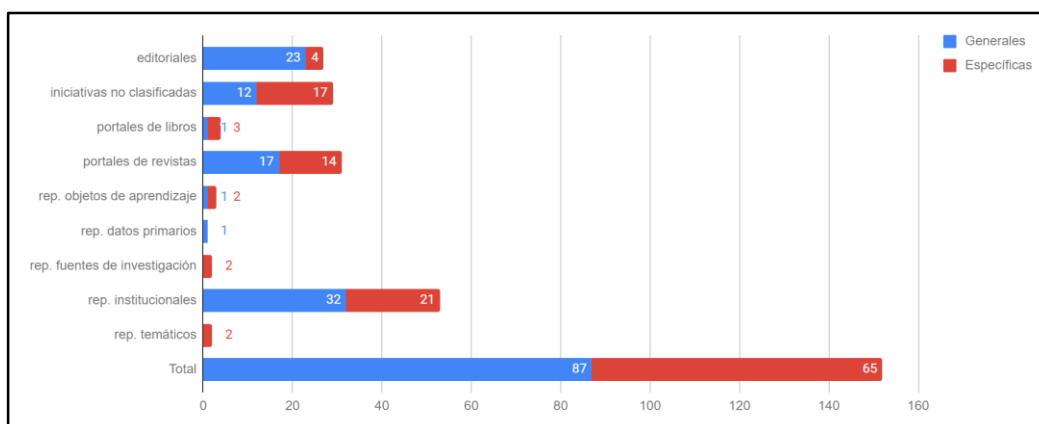


Gráfico 2: Tipo de iniciativas detectadas según área de la que depende su gestión.

Vinculado con lo anterior, en la tabla 3 se muestra la concentración de iniciativas por universidad, donde curiosamente se observan casos en los que se encontraron hasta 16 iniciativas diferentes por institución. Además de las diferencias que se originan en el tamaño de las universidades, esto muestra el interés, la diversidad y la dispersión que existe dentro del ámbito universitario nacional en cuestiones relacionadas a la difusión de la producción científica.

Iniciativas AA	UUNN	%	Universidades
Entre 11 y 16	4	6%	UNLP, UNMdP, UBA, UNC
Entre 4 y 6	7	11%	UNR, UnCOMA, UnaM, UNSAM, UNSa, UNPA, UNICEN
Entre 2 y 3	22	34%	UNS, UNRC, UNNE, UNLPAM, UNL, UNER, UNDEF, UNCuyo, UTN, UNT, UNSL, UNRN, UNRaf, UNQ, UNPSJB, UNM, UNLaR, UNLAM, UNLa, UNJu, UNPE, UNCA
Solo 1	13	20%	UNVM, UNTDF, UNSe, UNPAZ, UNLu, UNdeC, UNDAV, UNAJ, UNA, UADER, IUGNA, UPSO, UNGS
Sin iniciativas	19	29%	UNNOBA, UNSADA, UPC, IUNMa, IUPFA, IUSM, UNTreF, UnaB, UNLZ, UNO, UNPAZ, UNSO, UNAU, UNCAUS, UNF, UNLC, UNViMe, IUPA, UNAHUR,
Total UUNN	65	100%	

Tabla 3: Cantidad de iniciativas por UUNN

Caracterización de los repositorios institucionales

Como se mencionó anteriormente, durante el presente relevamiento en las 65 UUNN argentinas se localizaron un total de 61 repositorios, de los cuales 53 son institucionales, 3 de aprendizaje, 2 temáticos, 2 de fuentes y 1 específico de datos.

Teniendo en cuenta que el estudio anterior (Fushimi, 2018) había encontrado 46 repositorios institucionales contra 53 hallados ahora, el crecimiento del período 2015-2019 fue del 15%.

En el gráfico 3 se expone la distribución regional de los 53 RIs, en relación con la cantidad de universidades existentes, a la vez que se diferencia entre los generales y los específicos. Al igual que en el caso de la totalidad de las iniciativas analizadas, se observa una mayor concentración de repositorios en las regiones metropolitana, bonaerense y centro, en donde también el número de instituciones de educación superior es más elevado.

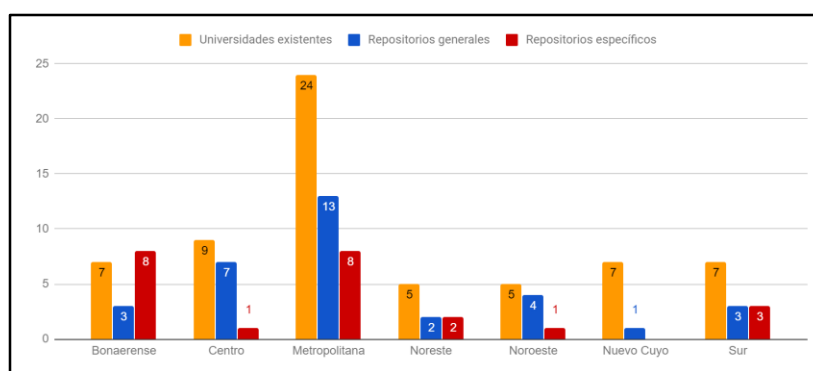


Gráfico 3. Distribución regional de las UUNN y de los RIs.

Cabe señalar, asimismo, que los 53 repositorios encontrados pertenecen a 27 de las 65 UUNN relevadas (41,5%), hecho que puede deberse a que varias de estas universidades son de reciente creación y es probable que aún no hayan avanzado en la organización de su RI ni en otro tipo de iniciativa de AA a su producción científica. Además, corresponde indicar que una de las universidades observadas posee únicamente repositorios específicos (como es el caso de UNMdP), lo que estaría mostrando la existencia de diferentes modelos de desarrollo en las universidades nacionales.

Por otra parte, respecto del estado de los repositorios institucionales, se encontró que el 87% estaba en funcionamiento (n=46), el 11% en desarrollo (n=6) y sólo 2% detenidos (n=1). Y en cuanto al software empleado para la gestión del RI, si bien se observó cierta diversidad en las soluciones implementadas por las distintas instituciones, es DSpace el más utilizado (29 repositorios -20 generales y 9 específicos-), seguido por Greenstone (11 repositorios -5 generales y 6 específicos-). Los restantes RIs emplean Omeka (2), Eprints (1), PMB (1) y desarrollos propios (2).

Grado de desarrollo de los repositorios institucionales

Para conocer el grado de desarrollo de los RIs de las UUNN argentinas, y teniendo en cuenta los indicadores propuestos por Fushimi (2018) -tal como se mencionó en la metodología-, se observó el *tamaño* en cuanto a cantidad de objetos digitales disponibilizados, la *visibilidad* a través de diversas fuentes, la existencia de *políticas* referenciadas en las webs de los propios repositorios, el año de puesta en funcionamiento para calcular la *antigüedad*, y su adhesión al SNRD.

El *tamaño* es una medida fundamental para evaluar el grado de desarrollo de una iniciativa, ya que un repositorio vacío o con escasos documentos no estaría cumpliendo adecuadamente su función. Sin embargo, para determinar esta variable es preciso considerar, además, las características de la universidad, su antigüedad y la cantidad de producción científico-académica que se genera en su ámbito, datos que generalmente no son de fácil acceso.

A partir de este relevamiento, y tal como se observa en el gráfico 4, se encontró que el 43% de los RIs (n=23) tiene más de 1.000 documentos depositados, lo cual representa una cantidad importante, e inclusive se halló que el 11% (n=5) tienen más de 10.000. Como dato referencial, la media de los valores de tamaño de los repositorios fue de 5.695, y la mediana de 978.

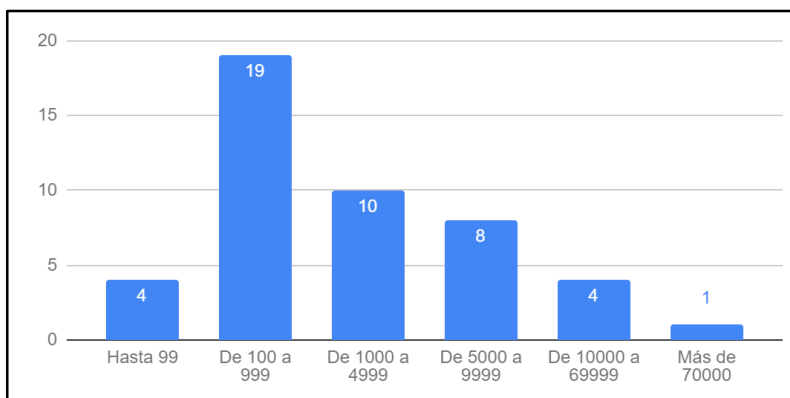


Gráfico 4. Cantidad de RIs agrupados según tamaño de sus colecciones

A fin de contextualizar el dato en análisis, se cruzó la variable de tamaño con el año de puesta en funcionamiento del repositorio (gráfico 5). Si bien la correlación es baja ($R^2=0,3765$) y existe dispersión, pudo observarse que los repositorios con mayor antigüedad son aquellos que mayor tamaño presentan, y viceversa.

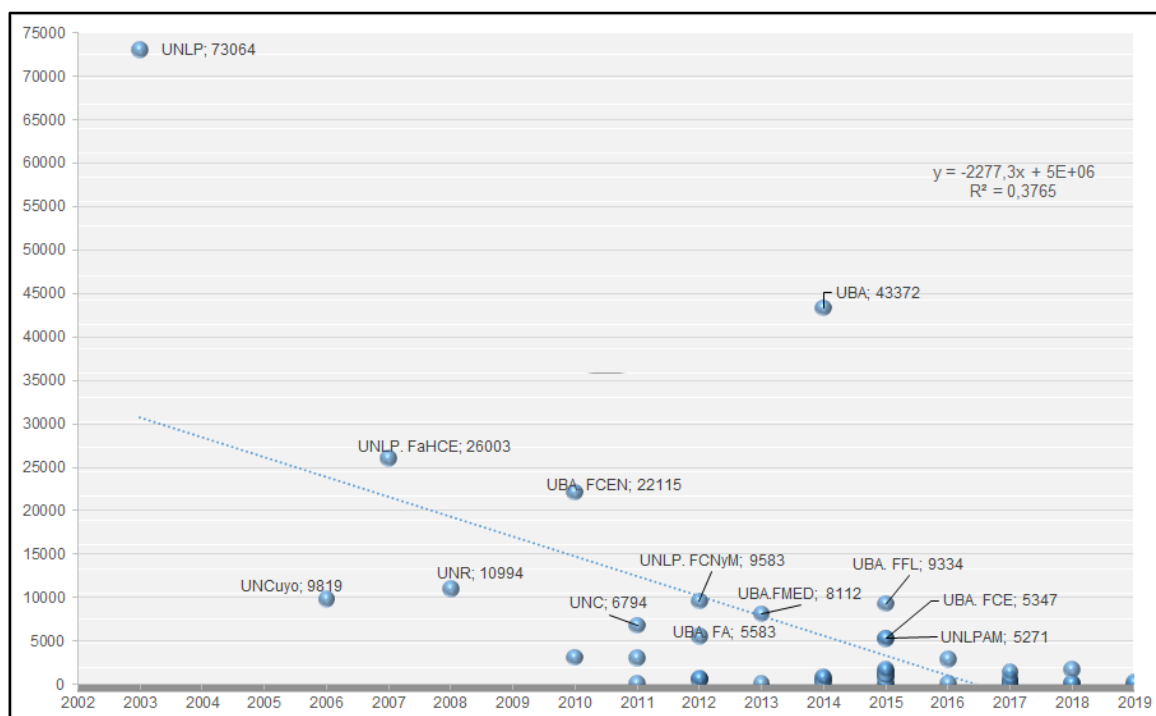


Gráfico 5. Tamanho versus antigüedad de los RIs.

En relación con la *visibilidad* de los repositorios, en el gráfico 6 se muestra la cantidad de RIs incluidos en cada una de las fuentes consultadas, donde la preponderancia de la BECYT y el SNRD por sobre el resto de las fuentes pone en evidencia el interés primordial de los gestores en ser incluidos en el Sistema Nacional. Asimismo, se encontró que un 30,4% (n=14) de los repositorios no están incluidos en ninguna de las fuentes, aunque el 41,3 % (n=19) está presente en cuatro y cinco fuentes.

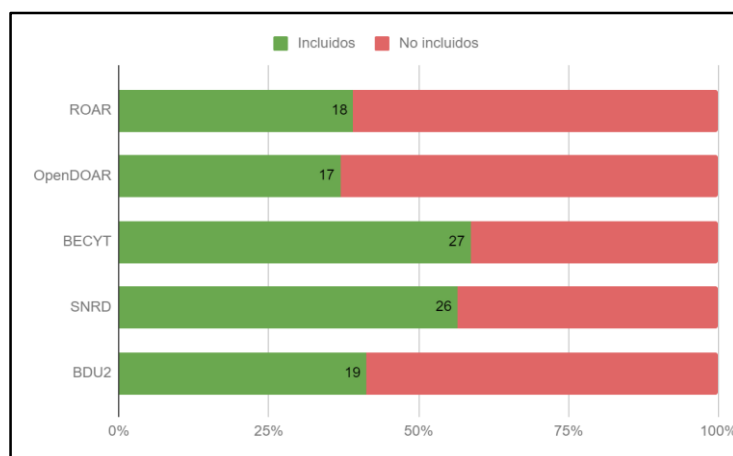


Gráfico 6. Visibilidade de los repositorios según fuente consultada.

En cuanto a las *políticas*, se observó que el 72% (n=33) de los RIs en funcionamiento cuenta con una normativa que aprueba su creación y fija criterios de organización. Muchas de estas políticas -en particular aquellas aprobadas con posterioridad a diciembre de 2013- se fundamentan en la ley 26.899 y sus considerandos. Por lo general tienen forma de resoluciones, reglamentaciones u ordenanzas, y han sido aprobadas por el órgano máximo de gobierno de las universidades que las poseen.

Respecto de la variable *antigüedad*, el gráfico 7 dispone la cantidad de repositorios puestos en funcionamiento por año desde 2003 hasta 2019, e incluye una línea de tendencia lineal positiva. Además, es posible observar que, de 2003 hasta 2013 -año de aprobación de la ley 26.899-, se crearon y pusieron en funcionamiento 16 repositorios institucionales, mientras que de 2014 a 2019 se desarrollaron los restantes 30, lo cual muestra, en cierta forma, la influencia de la ley en este segundo período.

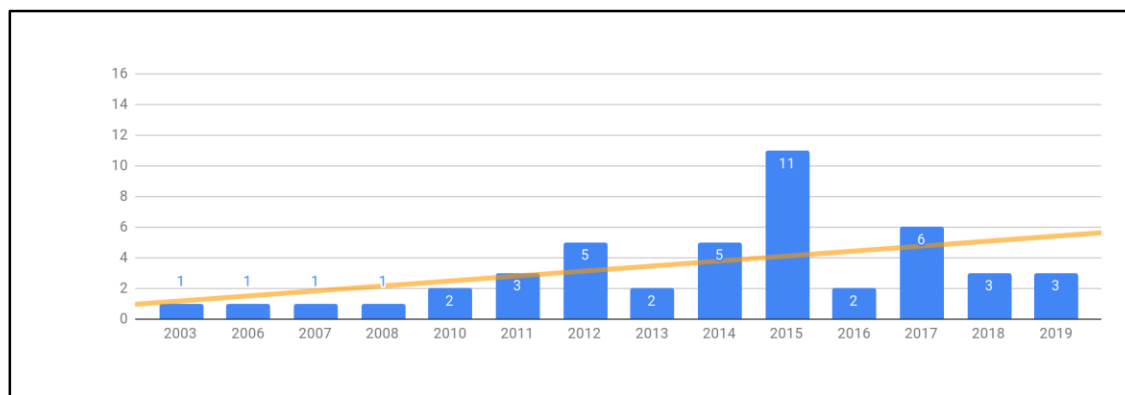


Gráfico 7. Repositorios puestos en funcionamiento por año durante 2003-2019.

Finalmente, en relación a la *adhesión al SNRD*, se pudo comprobar que de los 46 repositorios institucionales en funcionamiento, el 72% (n=33) estaba adherido, mientras que el restante 28% (n=13) aún no había tramitado su vinculación.

Grupos relacionados con la gestión

Para avanzar en el análisis de los grupos sociales involucrados y teniendo como base las categorías planteadas por Guédon (2009), se analizaron los datos relevados contemplando las particularidades del universo estudiado y las estructuras que suelen estar presentes en los organigramas de las universidades. Se utilizó, entonces, la siguiente categorización para identificar a los grupos involucrados en la gestión de los repositorios:

- Editoriales: Área de gestión editorial y/o editorial.
- Investigación: Área de gestión de investigación, posgrado y/o ciencia y técnica.
- Bibliotecas: Área de servicios de información bibliotecarios.
- Académica: Área de gestión académica.
- Tecnologías: Área de desarrollos tecnológicos, informática, educación a distancia, innovación educativa y campus virtuales.
- Organismos específicos: Área de gestión específica para el acceso abierto y/o la iniciativa en cuestión.

La categorización es compleja de analizar, ya que en muchos casos intervienen grupos sociales diversos cuya organización dentro de las instituciones no es homogénea. Por otra parte, y tratándose de repositorios de UUNN, no se han identificado entre los actores a las agencias de financiamiento, que sí se mencionan en trabajos como los de Guédon (2009) y en De Freitas y Lima Leite (2019).

Tras el análisis se encontró que el 70% (n=32) de los repositorios en funcionamiento están siendo gestionados actualmente por las bibliotecas de las instituciones, mientras que hay 10 RIs que son gestionados por más de un área (7 iniciativas por dos áreas y 3 repositorios por tres áreas).

Gestores	Repositorios generales	Repositorios específicos	Total	%
Biblioteca	19	13	32	70%
Investigación	7	2	9	20%
Organismo específico	5	1	6	13%
Académica	2	1	3	7%
Tecnologías	4	1	5	11%
Editoriales		1	1	2%
No identificado	2	1	3	7%

Tabla 4. Gestores de los RIs detectados

La relación de los investigadores con los repositorios institucionales

La vía verde del AA implica una actitud proactiva por parte de los autores para disponer una copia de su trabajo de forma abierta. Como sostienen Carr, Swan, y Harnad (2011, p. 108), en la era post-Gutenberg el lema debería ser “autoarchivar para prosperar”. No obstante, se sabe que los investigadores no lo han vivenciado de esa manera. Los estados nacionales, las instituciones y las comunidades han invertido tiempo, recursos y esfuerzos desarrollando repositorios como infraestructuras necesarias para alcanzar el acceso a la producción científica pero los índices de autoarchivo no han sido los esperados. Este hecho ha impulsado numerosas investigaciones en el campo sobre los hábitos de los investigadores en relación con el acceso abierto y el uso de repositorios a fin de determinar cómo estos artefactos se instalan en el ecosistema de la comunicación científica (Troll Covey, 2011; Kim, 2011; Fry, J., Spezi, V., Proberts, S., y Creaser, C., 2016; Serrano-Vicente, Melero, y Abadal, 2016, Borrego, 2017).

Para realizar un primer acercamiento a las actitudes de los investigadores argentinos con respecto a los RIs de las UUNN, se revisó la literatura publicada sobre el tema y se encontraron varios trabajos, los cuales han sido organizados en tres grupos para su análisis. En el primero se sitúan tres trabajos que abordan los hábitos de publicación y difusión en acceso abierto y los ponen en relación con los repositorios. El de Bongiovani, Guarnieri, Babini y López (2014) toma como universo la Universidad Nacional de Rosario (UNR) y analiza las opiniones y las prácticas de sus docentes e investigadores en cuanto a la difusión de su producción en abierto y sus necesidades en torno a la utilización del repositorio institucional. En la encuesta que realizaron estos autores encontraron que un 80% de los investigadores estaban de acuerdo con difundir sus trabajos en acceso abierto, y que sólo un 13% tenía por hábito autoarchivar. Entre las razones que detectaron está la falta de conocimiento del recurso (51,2%), tanto para depositar su trabajo como por la desmotivación por su potencialidad, es decir, que tienen baja expectativa respecto a que otros colegas la usen como herramienta para el descubrimiento, siendo este aspecto a su vez lo que más los motivaría a depositar (76%). También señalaron que los investigadores de Ciencias Sociales y Humanas están entre los que mayor uso hacen del repositorio junto con los de Informática, Comunicación y Electrónica. Asimismo, detectaron que los canales de difusión en abierto más usados son las revistas (36%), y los sitios webs de las facultades (33%), los sitios web de grupos o centros (27%) y personales (23%); los RIs no superaron el 20% y los temáticos el 5%.

El segundo trabajo a mencionar es el de San Martín, Bongiovani, Casali y Deco (2015), quienes encuestaron a profesores de la Universidad Nacional de Rosario, la Universidad Nacional del Litoral, la Universidad Tecnológica Nacional (Distritos de Rosario y Santa Fe), y la Universidad Nacional de Entre Ríos, en este caso para analizar necesidades y prácticas de difusión de trabajos en acceso abierto

y avanzar en el mejoramiento de una interfaz de autoarchivo. Como resultados hallaron que sólo un 15% había autoarchivado trabajos, un 30% no lo había hecho, un 44% desconocía la existencia de los repositorios y el 9% decía no tener RI para hacerlo. En cuanto a las motivaciones para autoarchivar, encontraron como principal el uso frecuente por parte de los profesores, estudiantes y público en general (73%), seguido por la mayor visibilidad de sus obras (52%) y la facilidad del mecanismo de autoarchivo (un 34%).

Por último, cabe señalar la investigación de Solaro y Lamberti (2015) en la que indagaron acerca de las prácticas, las percepciones y las valoraciones respecto a la publicación científica en abierto que tienen docentes-investigadores de la Facultad de Filosofía y Humanidades de la Universidad Nacional de Córdoba, en la que incluyeron aspectos relacionados con los repositorios y su uso. En las preguntas referidas al repositorio observaron que el 61% conocía acerca de la existencia del repositorio, pero sólo el 38% había autoarchivado. Los que no habían depositado respondieron que desconocían los procedimientos, carecían de tiempo y hábito, había falta de precisión en la normativa respecto a derechos de autor, priorizaban el modelo de publicación, entre otros. Al preguntar sobre el uso del repositorio como fuente de información, encontraron que la frecuencia de uso era baja ya que los encuestados percibían que sus contenidos no eran muy específicos.

En el segundo grupo se incluyeron dos trabajos que estudian a los repositorios como espacio de difusión de las producciones en relación con redes sociales académicas. El primero, de Miguel, González y Ortiz-Jaureguizar (2018), analiza las prácticas y preferencias de docentes-investigadores de tres Facultades diferentes de la Universidad Nacional de La Plata (UNLP) respecto al uso de ResearchGate (RG) frente a los RIs para socializar sus producciones. Detectaron que, si bien el nivel de presencia en los repositorios es mayor en general, la media de documentos disponibilizados en RG es más alta, hecho que -argumentaron- podría deberse a que los investigadores desconocen la existencia de los repositorios y son los bibliotecarios quienes depositan o cosechan la producción, sumado a que la práctica individual en la red social puede resultar más atractiva por tener procedimientos más ágiles, la comunicación que realizan y las métricas que proveen. Con su estudio, los autores afirman la hipótesis de las diferencias que ambos espacios ofrecen para la difusión: mientras que los repositorios se vinculan a lo institucional y la pertenencia de los investigadores a la institución, la red social RG responde a lógicas de comunidad disciplinar. El segundo trabajo a mencionar es el de Unzurrunzaga (2017), que toma como universo a los investigadores del Instituto de Investigaciones en Humanidades y Ciencias Sociales (CONICET-UNLP) respecto al uso de Academia y RG y las motivaciones del autoarchivo en éstas frente a las del repositorio institucional. La autora encontró que las principales diferencias entre los motivos de uso de ambas tecnologías residen en los mecanismos de autoarchivo (la facilidad de carga de las redes frente a complejidad-

burocracia), el procurar la difusión inmediata entre un público lector específico con interés potencial y la interconexión con colegas, funcionalidad típica de las redes sociales.

En el tercer y último grupo, situamos un solo trabajo en el que se relaciona el tema de acceso abierto y el sistema de evaluación científica nacional. Se trata del texto de Bongiovani, Miguel y Hernández-Pérez (2017), que describe las actitudes y opiniones de los investigadores desde su rol de evaluadores de pares. En este caso se entrevistaron a los investigadores que integran el Banco de Evaluadores del Ministerio de Educación y Deportes de la Nación, y que son quienes evalúan la carrera docente en las universidades nacionales a través del Programa de Incentivos a Docentes Investigadores. Se detectó que tienen hábitos similares a la hora de elegir una publicación y evaluar luego a sus pares; mantienen criterios tradicionales y pocos son quienes tienen en cuenta el acceso como un factor clave en la publicación y evaluación. En relación a la posibilidad de incluir como parámetro de evaluación la difusión de una copia de la producción a través de los repositorios, un 82% de los investigadores se mostró a favor (27% obligatorio, 55% recomendado).

Discusión y conclusiones

Como resultado de esta primera etapa de la investigación podemos decir que se constata el impacto de la Ley 26.899 en el incremento de la cantidad de repositorios producidos en las UUNN desde su sanción en 2013. Antes de la ley se habían creado 16, y de 2014 a 2019 se pusieron en funcionamiento otros 30, a lo que se suma el dato de que el 72% de ellos está adherido al SNRD. A la vez, se observó un predominio de RIs por sobre las demás tipologías de repositorios, hecho que resulta indicativo de la dificultad de las instituciones para organizar y mantener repositorios de datos primarios de investigación, también obligados por la legislación.

Por otra parte, se destaca que este crecimiento estuvo acompañado del surgimiento de otras iniciativas de publicación y difusión abierta de producción científica mediante la web, que no utilizan las vías tradicionales (revistas y repositorios) y no siguen los parámetros técnicos del AA, pero expresa el interés de las instituciones -y sus comunidades- por difundir lo que producen. Heterogeneidad y dispersión parecen ser las características dominantes del importante crecimiento de estas iniciativas de acceso abierto en las UUNN.

En cuanto al desarrollo de los RIs, se observó un importante crecimiento en su tamaño, medido en función de la cantidad de objetos digitales depositados, en particular en los más antiguos. Y también un enfoque en la visibilidad nacional, más que en la internacional, debido al bajo registro encontrado en los directorios o portales internacionales del rubro. Asimismo, en coincidencia con los estudios previos, se encontró que las bibliotecas son los actores que continúan liderando la

gestión de los RIs en un porcentaje alto, aunque también se halló la presencia de otras áreas dentro de las instituciones (áreas de investigación y tecnologías mayormente).

Por otro lado, la revisión de la literatura acerca de estudios previos sobre las actitudes de los investigadores frente a estos artefactos nos permitió tener un primer acercamiento a estos actores y arriesgar la generalización de que existe una comunidad científica reticente al depósito, con prácticas de comunicación y difusión alejadas de los repositorios y -agregamos en base a nuestra experiencia como gestoras de repositorios y docentes-investigadoras de universidades- con cierto desconocimiento de los requerimientos de la ley y desconcierto ante un sistema de evaluación de la investigación que no se condice con la legislación de acceso abierto nacional y que promueve la circulación en sistemas comerciales de grandes empresas frente a la disponibilidad en abierto.

Ante este panorama, se considera necesario repensar los repositorios en tanto artefactos sociotécnicos complejos, con una mirada situada principalmente en la comunidad de origen. La interoperabilidad e integración de todos los sistemas existentes dentro del circuito de la comunicación académica son, en este sentido, fundamentales para lograr su éxito y que los investigadores de las instituciones jerarquicen el espacio, más allá de la existencia de mandatos para la difusión de su producción, frente a plataformas comerciales que ofrecen servicios y desarrollo más ágiles y efectivos. Está claro que los RIs universitarios deben mejorar y evolucionar con servicios adaptados a las necesidades de los investigadores, que hagan efectivo e inmediato el acceso abierto a su producción y datos. Y principalmente, que la comunidad académica se apropie del conocimiento que generó con recursos públicos, disponibilizándolo para toda persona interesada.

Referencias

- Babini, D. (2019). La comunicación científica en América Latina es abierta, colaborativa y no comercial. Desafíos para las revistas. *Palabra Clave (La Plata)*, 8(2), e065. <https://doi.org/10.24215/18539912e065>
- Bongiovani, P., Guarnieri, G., Babini, D., y López, F. (2014). Acceso abierto en la Universidad Nacional de Rosario. Necesidades y prácticas de los docentes/investigadores. *Información, cultura y sociedad*, 30, 13-33. Recuperado de <http://ppct.caicyt.gov.ar/index.php/ics/article/view/4591>
- Bongiovani, P., Miguel, S., y Hernández-Pérez, T. (2017). Actitudes y percepciones de los evaluadores de la carrera científica en Argentina sobre la publicación en acceso abierto. *Revista española de Documentación Científica*, 40(2), e171. <http://dx.doi.org/10.3989/redc.2017.2.1404>

- Bongiovani, P. y Nakano, S. (2011). Acceso Abierto en Argentina: La experiencia de articulación y coordinación institucional de los repositorios digitales en ciencia y tecnología". *e-colabora, Revista de ciencia, educación, innovación y cultura apoyadas por redes de tecnología avanzada*, 1(2), 163-179. Recuperado de [https://www.researchgate.net/publication/233927683 Acceso Abierto en Argentina La experiencia de articulacion y coordinacion institucional de los repositorios digitales en ciencia y tecnologia](https://www.researchgate.net/publication/233927683_Acceso_Abierto_en_Argentina_La_experiencia_de_articulacion_y_coordinacion_institucional_de_los_repositorios_digitales_en_ciencia_y_tecnologia)
- Borrego, A. (2017). Institutional repositories versus ResearchGate: The depositing habits of Spanish researchers. *Learned Publishing*, 30(3), 185-192. <http://dx.doi.org/10.1002/leap.1099>
- Carr, L., Swan, A. y Harnad, S. (2011). Creación y mantenimiento del conocimiento compartido: contribución de la University of Southampton. *El profesional de la Información*, 20(1), 102-110. Recuperado de <http://www.elprofesionaldelainformacion.com/contenidos/2011/enero/13.pdf>
- de Freitas, M.A; Lima Leite, F.C. (2019) Atores do sistema de comunicacao científica: apontamentos para discussao de suas funcoes. *Inf. Inf.*, 24(1), 273-299.
- De Volder, C. (2008). Los repositorios de acceso abierto en Argentina: situación actual. *Información, Cultura y Sociedad*, 19, 79-98. Recuperado de <http://revistascientificas.filo.uba.ar/index.php/ICS/article/view/83>
- Fry, J., Spezi, V., Proberts, S., y Creaser, C. (2016). Towards an understanding of the relationship between disciplinary research cultures and open access repository behaviors. *Journal of the Association for Information Science and Technology*, 67(11), 2710-2724. <https://doi.org/10.1002/asi.23621>
- Fushimi, M., y Banzato, G. (2010). *Las políticas de acceso abierto en las universidades estatales argentinas: un análisis a través de la web*. Ponencia presentada en VI Jornadas de Sociología de la UNLP, 9-10 diciembre 2010, UNLP, La Plata, Argentina. Recuperado de http://www.memoria.fahce.unlp.edu.ar/trab_eventos/ev.931/ev.931.pdf
- Fushimi, M. (2018) *Desarrollo de repositorios digitales institucionales en las universidades nacionales en Argentina, período 2004-2015* (Tesis de maestría). Universidad Nacional de Quilmes. Recuperado de <http://www.memoria.fahce.unlp.edu.ar/tesis/te.1629/te.1629.pdf>

- Guédon, J. C. (2009). It's a repository, it's a depository, it's an archive...: open access, digital collections and value. *ARBOR Ciencia, Pensamiento y Cultura*, 185(737), 581-595. <http://dx.doi.org/10.3989/arbor.2009.i737.315>
<http://dx.doi.org/10.5433/1981-8920.2019v24n1p273>
- Kim, J. (2011). Motivations of faculty self-archiving in institutional repositories. *The Journal of Academic Librarianship*, 37(3), 246-254. <https://doi.org/10.1016/j.acalib.2011.02.017>
- Larivière, V., y Sugimoto, C. R. (2018). Do authors comply when funders enforce open access to research? *Nature*, 562(7728), 483. <https://doi.org/10.1038/d41586-018-07101-w>
- Ley N° 26.899, 2013, Repositorios digitales institucionales de acceso abierto. *Boletín Oficial de la República Argentina*, 32.781, Primera Sección, 3-4. Recuperado de <http://servicios.infoleg.gob.ar/infolegInternet/anexos/220000-224999/223459/norma.htm>
- Miguel, S., González, C., y Ortiz-Jaureguizar, E. (2018). Preferencias de investigadores y prácticas institucionales/disciplinares en la difusión y socialización de los resultados de investigación. *Información, Cultura y Sociedad*, 38, 53-76. Recuperado de <http://revistascientificas.filo.uba.ar/index.php/ICS/article/view/3989>
- Nakano, S., y Azrilevich, P. A. (2017). *El acceso abierto y la implementación de la Ley 26.899 en la Argentina*. Presentado en VII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales de América Latina (BIREDIAL-ISTEC'17) y XII Simposio Internacional de Biblioteca Digitales (SIBD'17) (La Plata, 2017). Recuperado de <http://hdl.handle.net/10915/63553>
- Pené, M.G., Unzueta, C. y Borrell, M. (2016). *Repositorios institucionales universitarios argentinos, un acercamiento a sus colecciones*. Presentado en IV Jornadas de Intercambio y Reflexión acerca de la Investigación en Bibliotecología, La Plata, 2015. Recuperado de http://www.memoria.fahce.unlp.edu.ar/trab_eventos/ev.5293/ev.5293.pdf
- Piowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley A, West, J., y Haustein, S. (2018). The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*, 6, e4375. doi:[10.7717/peerj.4375](https://doi.org/10.7717/peerj.4375)

- San Martín, P., Bongiovani, P., Casali, A., & Deco, C. (2015). Study on Perspectives Regarding Deposit on Open Access Repositories in the Context of Public Universities in the Central-Eastern Region of Argentina. *Scholarly and Research Communication*, 6(1). <http://dx.doi.org/10.22230/src.2015v6n1a145>
- Serrano-Vicente, R., Melero, R., y Abadal, E. (2016). Open Access Awareness and Perceptions in an Institutional Landscape. *The Journal of Academic Librarianship*, 42(5), 595-603. <https://doi.org/10.1016/j.acalib.2016.07.002>
- Solaro, G. N., y Lamberti, A. (2015). *Prácticas de publicación, percepciones y valoraciones en torno al Acceso Abierto de los investigadores de la Facultad de Filosofía y Humanidades de la Universidad Nacional de Córdoba* (Tesis de Licenciatura). Universidad Nacional de Córdoba, Córdoba (Argentina). Recuperado de <http://hdl.handle.net/10760/28942>
- Troll Covey, D. (2011). Recruiting Content for the Institutional Repository: The Barriers Exceed the Benefits. *Journal of Digital Information*, 12(3). Recuperado de <https://journals.tdl.org/jodi/index.php/jodi/article/view/2068>
- Unzurrunzaga, C. (2017). *Hábitos de los investigadores de Humanidades y Ciencias Sociales para visibilizar sus publicaciones en el entorno digital. Un estudio de caso argentino* (Trabajo final de Máster). Universidad de Salamanca. Recuperado de <https://gredos.usal.es/jspui/handle/10366/138126>

DIA 2

PUBLICAÇÕES ELETRÔNICAS, AVALIAÇÕES E MÉTRICAS

Impacto na Ciência

Visibilidade

Preservação digital de periódicos eletrônicos

Avaliação ergonômica de repositórios digitais institucionais – Revisão integrativa da literatura

Wilma Honório dos Santos¹, Prof. Dr. Felipe Mancini², Prof. Dr. Luciano Gamez³

¹ Mestre em Ciências. Universidade Federal de São Paulo (Unifesp). Email: wilma.santos@unifesp.br

² Gestão de Informática em Saúde. Universidade Federal de São Paulo (Unifesp)/ Universidade Aberta do Brasil (UAB). Email: fmancini@unifesp.br

³ Universidade Federal de São Paulo (Unifesp)/ Universidade Aberta do Brasil (UAB). Email: lucianogamez@unifesp.br

Resumo

Repositórios digitais são ambientes informacionais para gerenciamento e controle da produção científica e acadêmica de instituições e/ou comunidades. Oferecem vantagens como acesso irrestrito, interoperabilidade dos dados e preservação da informação em longo prazo. Entretanto, eles podem apresentar lacunas como falhas de navegação, baixa usabilidade e acessibilidade, buscas limitadas, pouca divulgação do ambiente e pouca ou nenhuma utilização de serviços personalizáveis. A partir desse contexto, torna-se necessária uma avaliação ergonômica destes. Este trabalho teve como objetivo realizar uma revisão integrativa da literatura para avaliar técnicas de usabilidade aplicadas à ergonomia de RDIs. A revisão integrativa da literatura buscou para avaliar o estado da arte das técnicas de usabilidade aplicadas à ergonomia de repositórios digitais institucionais e os métodos aplicados para sua avaliação, o fichamento dos artigos incluídos na RIL para auxiliar na escolha do método de avaliação de usabilidade, seleção de um conjunto de requisitos para inspeção de usabilidade em RDIs. A estratégia da revisão integrativa da literatura nas bases de dados utilizadas e descritores usados permitiu responder as perguntas de pesquisa e mostrou que os testes aplicados melhoram a usabilidade dos RDs. Mostrou ainda que os modelos, métodos e técnicas utilizados para avaliação de usabilidade são o teste de cenários de interação com tarefas e teste de inspeção por *checklist*. E, ainda que os métodos de avaliação de usabilidade aprimoram a experiência do usuário quanto ao uso da interface.

Palavras-chave: repositórios digitais, repositórios digitais institucionais, ergonomia, usabilidade.

Introdução

A internet como ferramenta de suporte educacional e sua utilização apoia o processo de ensino-aprendizagem. Associada às Tecnologias Digitais da Informação e Comunicação (TIDC) possibilitam a criação de documentos digitais (DD), responsáveis pela preservação do patrimônio cultural, possibilitam a disseminação de informações e fortalecem a construção do conhecimento.^[1]

Os repositórios digitais (RDs) surgiram para assegurar a preservação e acesso desses DD. Estes possibilitam armazenar, organizar, gerir e acessar a produção científica e acadêmica, gerenciar as comunidades e suas coleções. Proporcionam visibilidade para a instituição, interoperabilidade de dados, controle e armazenamento da produção dos pesquisadores, preservação da informação em longo prazo, auto arquivamento, acesso livre e diminuição de custos de publicação.^[1]

As instituições acadêmicas utilizam os RDIs para dar suporte e gerenciar a informação científica, principalmente das atividades de pesquisa e ensino. Eles são utilizados para melhorar a comunicação interna e externa da instituição; maximizar a acessibilidade; a visibilidade e o impacto da produção científica da instituição; retroalimentar a atividade de pesquisa e apoiar os processos de ensino e aprendizagem; respaldar as publicações da instituição; contribuir para a preservação dos conteúdos digitais ou acadêmicos produzidos pela instituição ou seus membros acesso livre e diminuição de custos de publicação.^[1]

Para a construção de um RD devem ser considerados requisitos de navegação, usabilidade e acessibilidade, buscas, divulgação do ambiente e serviços personalizáveis. Neste contexto os RDs devem ser avaliados no que tange à ergonomia e usabilidade para proporcionar uma interação efetiva entre o usuário e o material disponível em um RD e a sua interface.^[1,4]

Objetivo

O objetivo deste trabalho é realizar uma revisão integrativa da literatura para avaliar as técnicas de usabilidade aplicadas à ergonomia de RDs e RDIs.

Referencial teórico

Repositórios digitais (RD) e repositórios digitais institucionais (RDIs)

Os RDs podem ser: Temáticos (RDT): quando focam uma determinada área do conhecimento ou Institucionais (RDIs): quando pertencem a uma determinada instituição e centrados na divulgação da produção científica local, trabalhos de pesquisa e ensino de comunidades acadêmicas, teses, dissertações etc.^[1,3]

Os RDIs podem ser vistos são ferramentas aplicadas à gestão do conhecimento científico. Potencializam a condução de processos para a criação, o compartilhamento, a disseminação e o uso do conhecimento.^[1]

Com o aumento da produção de informação em formato de DD, torna-se importante a garantia da sua disponibilização e preservação ao longo do tempo. Essa preocupação envolve tanto os produtores dos dados quanto os órgãos detentores dessa informação.^[5]

Neste contexto surgiram os RDs, que são ambientes informacionais para o armazenamento e gerenciamento de DD que permitem a organização e o acesso da produção científica e acadêmica, o gerenciamento de comunidades e coleções científicas e facilitam a implementação das políticas e estratégias de preservação.^[6,7]

Ao RD compete o compromisso de preservar, gerenciar e possibilitar o acesso contínuo em longo prazo à documentos arquivísticos digitais autênticos.^[6] Segundo o Conselho Nacional de Arquivos^[6] (CONARQ), os RDs devem gerenciar os documentos e metadados de acordo com as práticas e normas da Arquivologia, especificamente relacionadas à gestão documental, descrição arquivística multinível e preservação, e proteger as características do documento arquivístico, em especial a autenticidade (identidade e integridade) e a relação orgânica entre os documentos.

Entre as vantagens dos RDIs temos que estes proporcionam visibilidade para a instituição, interoperabilidade de dados, controle e armazenamento da produção científica, preservação da informação em longo prazo, auto arquivamento, acesso livre, diminuição de custos de publicação.^[2]

Ergonomia

O termo ergonomia significa o estudo das leis do trabalho e pode ser definido como o estudo científico da relação entre o homem e seus meios, métodos e espaços de trabalho e ainda, como os sentidos e as capacidades motoras permitem às pessoas utilizarem máquinas e ferramentas.^[8,9]

Seu objetivo é elaborar, mediante a contribuição de diversas disciplinas científicas que a compõem, um corpo de conhecimentos que, dentro de uma perspectiva de aplicação, deve resultar em uma melhor adaptação ao homem dos meios tecnológicos, dos ambientes de trabalho e de vida.^[9,10]

Ergonomia cognitiva

Com a ampla utilização dos computadores, o estudo da ergonomia foi ampliado para analisar a capacidade mental que possibilita às pessoas produzirem, recuperarem e compreenderem informações geradas pelas TDIC e fez surgir a ergonomia cognitiva.^[8,11]

A ergonomia cognitiva preocupa-se com os aspectos da atividade mental realizada pelo usuário numa determinada atividade e busca otimizar o esforço despendido para compreender e desenvolver a tarefa, como também facilitar o processo mental para a tomada de decisões e execução de determinada ação.^[10-12]

Carga mental de trabalho

Todos os elementos da interface devem reduzir a carga cognitiva e perceptiva do usuário, e aumentar da eficiência do diálogo. Assim, quanto maior for a carga mental de trabalho, maior será a probabilidade de cometer erros e quanto menos ações forem necessárias, mais rápidas serão as interações.^[8]

Usabilidade

Segundo Nielsen (1994), a usabilidade é um requisito de qualidade de *software* necessário e exigido para atingir a qualidade de um sistema computacional permitindo que ele seja utilizável e de fácil aprendizado. Um sistema que tenha boas taxas de usabilidade permite que seus usuários o utilizem de forma satisfatória, agradável e produtiva e assim, atinja seu objetivo.^[13]

Uma interface que possui boa usabilidade evita que seu usuário tenha que aprender procedimentos complexos, ajuda na memorização das atividades no sistema, orienta na exploração de seu conteúdo, protege contra erros e facilita procedimentos e reduz a carga física e mental do usuário, além da diminuir o tempo gasto para realizar uma tarefa.^[11]

Repositórios, ergonomia cognitiva e usabilidade

Para a obtenção de uma interação efetiva entre o usuário e o material disponível em um RD a sua interface deve agregar conceitos de usabilidade, que é a qualidade relacionada à facilidade de uso e de aprendizado, e de ergonomia, que é a qualidade de adaptação de uma interface ao perfil do usuário.^[2,7,8,12,14-18]

A usabilidade dos RDs são fundamentais por serem considerados fatores relevantes no ambiente informacional, com vistas a proporcionar ao usuário que acessa o sistema e com ele interage uma boa navegabilidade.^[19]

Um alto grau de usabilidade de uma interface reflete em usuários realizando tarefas com facilidade, rapidez e satisfação.^[4,11,10,13,12]

Estudos realizados em RDs mostram que eles podem possuir lacunas como falhas de navegação, baixa usabilidade, buscas limitadas, pouca divulgação do ambiente e pouca ou nenhuma utilização de serviços personalizáveis.^[2,7,14,15,20-22]

Avaliação de usabilidade

A avaliação da usabilidade pode ser realizada de duas formas: teste de usabilidade e inspeção de usabilidade.

O teste de usabilidade refere-se a atividades sistemáticas com objetivo de verificar como uma pessoa ou um grupo de pessoas interagem com a aplicação e como a mesma interfere positiva ou negativamente em suas atividades, ou seja,

tem como principal intenção verificar a capacidade de interação oferecida pela interface ao seu usuário.^[4,11,13,16]

São realizados na forma de cenários de interação onde o usuário segue tarefas específicas e pré-determinadas.^[3]

A inspeção de usabilidade refere-se às atividades que visam verificar se uma interface está de acordo com um determinado padrão de qualidade como os Critérios Ergonômicos de Dominique Scapin e Christian Batien ou as Heurísticas de Jakob Nielsen.^[12,13]

As principais vantagens da inspeção baseada em *guidelines* são as possibilidades de aplicação sem a necessidade de envolvimento de usuários, a rapidez e a facilidade de aplicação, podendo ser adotada, inclusive, por avaliadores não especialistas em usabilidade.^[23]

Materiais e métodos

Este trabalho trata-se de uma pesquisa qualitativa, com abordagem em estudo de caso para aplicar as técnicas de inspeção de usabilidade (preditiva).

Para atender aos preceitos éticos em pesquisas com seres humanos, conforme consta na Resolução nº 466/12 do Conselho Nacional de Saúde, este trabalho de pesquisa foi aprovado sob parecer de número 1.545.877 e CAAE 55570716.2.0000.5505, pelo Comitê de Ética em Pesquisa (CEP) do Setor de Ciências da Saúde da Universidade Federal de São Paulo (Unifesp).

Para alcançar os objetivos principal e específicos propostos neste trabalho, foi realizada uma revisão integrativa da literatura (RIL) com o objetivo de investigar o estado da arte das técnicas de usabilidade aplicadas à ergonomia de RDs e RDIs.

Revisão integrativa da literatura

A escolha da RIL surgiu devido a este tipo de revisão proporcionar uma análise ampla da literatura, contribuir para discussões sobre métodos e resultados de pesquisas, assim como reflexões sobre a realização de futuros estudos. O propósito inicial deste método de pesquisa é obter um profundo entendimento de um determinado fenômeno baseando-se em estudos anteriores, permite a combinação de dados da literatura empírica e teórica que podem ser direcionados à definição de conceitos, identificação de lacunas nas áreas de estudos, revisão de teorias e análise metodológica dos estudos sobre um determinado assunto.^[24]

Pergunta de Pesquisa

Para nortear a RIL foram criadas as seguintes perguntas de pesquisa:

Geral

- O uso de técnicas de usabilidade aponta melhorias para a ergonomia de RDIs?

Específicas

- Quais são os modelos, métodos e técnicas utilizados na avaliação de usabilidade de RDIs?
- Métodos de avaliação de usabilidade aprimoram a experiência do usuário quanto ao uso da interface?

CrITÉrios de inclusão

Os critérios de inclusão da RIL foram: (1) artigos que contenham técnicas e/ou métodos de avaliação ergonômica da usabilidade de RDs; (2) artigos descritivos relacionando a avaliação da usabilidade com aplicações em RDs e/ou RDIs.

CrITÉrios de exclusão

Para aperfeiçoar a objetividade e a validade da pesquisa, excluiu-se artigos que se enquadram em, pelo menos, um dos 5 critérios de exclusão: (1) artigos que não estejam escritos em português ou na língua inglesa; (2) artigos similares com resultados duplicados em diferentes bases de dados; (3) artigos incompletos (somente resumo, tutoriais, *whitepapers* ou *keynotes*); (4) capítulos de livros; (5) artigos que não contenham alguma aplicação relacionada ao contexto de usabilidade em RDs e RDIs.

Estratégia de Busca

Para obtenção dos artigos do estudo, construiu-se incrementalmente uma *string* de busca:

#1 USABILIDADE OR USABILITY OR (ERGONOMIA COGNITIVA) OR (COGNITIVE ERGONOMICS) OR (AVALIAÇÃO ERGONÔMICA) OR (ERGONOMIC EVALUATION)

#2 (MÉTODOS DE AVALIAÇÃO) OR (EVALUATION METHODS) OR (EVALUATION) OR (METHOD*) OR (MÉTODO)

#3 (REPOSIT* DIGITA*) OR (REPOSIT* INSTITUTIONAL*) OR (REPOSIT* DIGIT* INSTITUTIONAL*) OR (REPOSIT*)

#4 #1 AND #2 AND #3

A *string* de busca foi baseada nos termos da questão de pesquisa: usabilidade, ergonomia cognitiva e RDIs. O processo de pesquisa consistiu em uma busca automática, via *Portal de Periódicos CAPES/MEC*¹ usando a Rede Privada Virtual (VPN) da Unifesp, na *Web of Science*, pois esta base de dados indexa mais de 12.000 revistas de impacto em todo o mundo, incluindo revistas como as da *Association for Computing Machinery (ACM) Digital Library* ou as do *Institute of*

¹<http://www.periodicos.capes.gov.br/>

Electrical and Electronics Engineers (IEEE). A pesquisa foi realizada no dia 26/09/2016.

Artigos selecionados

Após a busca automática, os artigos resultantes foram: (1) pré-selecionados segundo o título e o *abstract*; (2) selecionados segundo os critérios de inclusão e exclusão para avaliação.

Avaliadores externos

Participaram desta RIL 5 pesquisadores que preencheram os seguintes critérios de inclusão: (1) graduado em nível superior; (2) cursando pós-graduação *stricto sensu*; (3) possuir conhecimento de Metodologia Científica e (4) apresentar disponibilidade para avaliação dos artigos.

Os artigos foram distribuídos aleatoriamente entre os pesquisadores, que após a leitura do material de apoio e dos artigos selecionados, responderam ao questionário de avaliação para compor a RIL apresentado no Quadro 1.

Material de apoio

Foi criado um blog “MITEADS - Métodos e Inovações Tecnológicas para Educação a Distância em Saúde² com o seguinte material de apoio para transmitir aos pesquisadores o conhecimento sobre o assunto: Conceitos de usabilidade, ergonomia, RDs e RDIs; Tutorial *Zotero*³; Relação completa das referências dos artigos a serem avaliados e, o formulário de avaliação dos artigos selecionados criado no *Google Forms*⁴. Utilizou-se o gerenciador de referências *Zotero*⁵ de maneira compartilhada com os pesquisadores a fim de organizar e disponibilizar os artigos selecionados e, auxiliar na composição das referências bibliográficas desta RIL, o gerenciador de planilhas *Microsoft Excel 2016* para a criação de tabelas dinâmicas e gráficos como ferramentas de apoio para melhorar a compreensão dos resultados.

Processo de avaliação

Os artigos que satisfizeram os critérios de inclusão contêm informação suficiente para identificar os modelos, métodos e técnicas utilizados para essa RIL e foram avaliados em forma de respostas a um questionário com questões pré-estabelecidas sobre o assunto, pertinentes à pergunta da pesquisa, algumas com

²<http://miteds.blogspot.com.br/>

³<https://www.zotero.org/support/pt/start>

⁴<https://docs.google.com/forms/u/0/>

⁵<https://www.zotero.org>

alternativas afirmativas ou negativas ou a escala de *Likert* conforme o Quadro 1 mostrado a seguir.

Quadro 1: Questionário da revisão integrativa da literatura

Perguntas	Alternativas
O(s) teste(s) aplicado(s) melhora(m) a usabilidade do RD?	Concordo totalmente Concordo parcialmente Não concordo parcialmente Não concordo totalmente Não se aplica
Foi proposto um novo método de teste de usabilidade por inspeção com o uso de <i>checklist</i> ?	Sim Não
Foi proposto um novo método de teste de avaliação de cenários de interação com tarefas?	Sim Não
Avalia a usabilidade pelo teste de inspeção por <i>checklist</i> ?	Sim Não
Avalia a usabilidade por teste avaliação de cenários de interação com tarefas?	Sim Não
Combinam as duas técnicas	Sim Não
Compara as duas técnicas?	Sim Não
Recomenda a combinação dessas técnicas?	Sim Não
Considera o teste de inspeção por <i>checklist</i> mais eficiente?	Sim Não
Considera o teste avaliação de cenários de interação com tarefas mais eficiente?	Sim Não
Recomenda a combinação dessas técnicas?	Sim Não
Aponta melhorias na interface gráfica com o teste de inspeção por <i>checklist</i> ?	Sim Não
Aponta melhorias na interface gráfica por avaliação por cenários de interação com tarefas?	Sim Não
Reimplementou a interface gráfica?	Sim Não
Avalia repositórios digitais institucionais (RDI)?	Sim Não

Fichamento dos artigos da revisão integrativa da literatura

Foram feitos pelos autores deste trabalho a leitura e o fichamento dos artigos incluídos na RIL para subsidiar na escolha do método de avaliação de usabilidade e seleção de um conjunto de requisitos para inspeção de usabilidade em RDIs. Criou-se uma tabela contendo: Título da Referência; Autor(es); Ano; Avaliação da usabilidade e o Resumo.

Resultados

Resultados da RIL

Na etapa da RIL foram utilizados 18 artigos que apresentaram os itens para responder à pergunta de pesquisa.

O Quadro 2 mostra a compilação dos resultados da RIL. O Quadro 3 mostra a referência, os autores, ano de publicação e o resumo dos artigos avaliados da RIL.

Quadro 2: Compilação dos resultados da RIL

Compilação dos resultados da RIL	%
Apontam melhorias à interface gráfica usando cenários de interação com tarefas	83
Apontam melhorias à interface gráfica com o uso de teste de inspeção por <i>checklist</i>	72
Recomendam a combinação das duas técnicas	56
Proposta de um novo método de teste de avaliação de cenários de interação com tarefas	50
Avaliação da usabilidade por teste de cenários de interação com tarefas	50
Avaliação da usabilidade com o uso de teste de inspeção por <i>checklist</i>	45
Consideram o teste de avaliação de cenários de interação com tarefas mais eficiente	44
Proposta de um novo método de teste de inspeção de usabilidade por <i>checklist</i>	39
Combinam as duas técnicas	33
Consideram o teste de inspeção por <i>checklist</i> mais eficiente	33
Os testes aplicados melhoram a usabilidade do RD	28
Comparação das duas técnicas	28

Fichamento dos artigos da revisão integrativa da literatura

A seguir, no Quadro 3, está apresentado o fichamento de todos artigos incluídos na RIL.

Quadro 3: Revisão integrativa da literatura (RIL)

Referência [número da referência]	Autores	Ano	Avaliação da usabilidade	Resumo
Uma estratégia de avaliação em repositórios digitais ^[12]	Camargo LSA, Vidotti SBG.	2008	Testes de inspeção por <i>checklist</i> com heurísticas	Propõe uma estratégia de avaliação da usabilidade de um RDI baseada em testes de inspeção por <i>checklist</i>
Uma proposta metodológica para avaliação de bibliotecas digitais: usabilidade e	Bohmerwald P.	2005	Testes de inspeção por <i>checklist</i> com grau de satisfação do usuário	Estudo de avaliação usabilidade em RDI. A coleta de dados foi composta por cinco etapas: (1) questionário



Referência [número da referência]	Autores	Ano	Avaliação da usabilidade	Resumo
comportamento de busca por informação na Biblioteca Digital da PUC-Minas ^[21]			Avaliação por cenários de interação com tarefas	sobre o contexto do usuário, (2) teste de usabilidade em laboratório, (3) questionário de satisfação, (4) estudo do comportamento de busca dos usuários pelo teste com busca livre e (5) análise do <i>log</i>
Repositórios institucionais: avaliação da usabilidade na Fundação Oswaldo Cruz (Fiocruz) ^[22]	Veiga VSO, Pimenta DN, Machado R, Silva A, Silva CH.	2013	Avaliação por cenários de interação com tarefas	Estudo da avaliação da usabilidade do RDI da Fiocruz baseada em cenários de interação com tarefas
Avaliação da usabilidade da interface do Lume Repositório Digital da UFRGS ^[25]	Santos DB	2013	Avaliação por cenários de interação com tarefas	Estudo de usabilidade em RDI do Lume Repositório Digital da UFRGS. Avaliação da usabilidade de um RDI baseada em cenários de interação com tarefas e teste de inspeção por <i>checklist</i>
Avaliação ergonômica de interfaces da SciELO–Scientific Electronic Library Online ^[26]	Oliveira ER.	2001	Avaliação por cenários de interação com tarefas Testes de inspeção por <i>checklist</i> com os Critérios Ergonômicos de Dominique Scapin e Christian Batien	Estudo de avaliação da usabilidade do RDI da SciELO baseada em cenários de interação com tarefas e teste de inspeção por <i>checklist</i>
<i>A study of usability of Web-based software repositories. In: Software Methods and Tools</i> ^[27]	Clayton N, Biddle R, Tempero E.	2000	Teste de inspeção por heurísticas Teste avaliação de usabilidade por carga de trabalho cognitivo.	Avaliação do RD por teste de inspeção por heurísticas e o teste avaliação de usabilidade por carga de trabalho cognitivo. Considera que o ponto fraco dos RDs é a usabilidade e recomenda a combinação das duas técnicas
<i>Usability study of digital institutional repositories</i> ^[28]	Hee Kim H, Ho Kim Y.	2008	Avaliação por cenários de interação com tarefas	Avaliação da usabilidade de um RDI baseada em



Referência [número da referência]	Autores	Ano	Avaliação da usabilidade	Resumo
				cenários de interação com tarefas
Análise da interação de usuários com repositórios institucionais de instituições federais de ensino superior brasileiras ^[29]	Marcelino C de S Siebra S de A.	2016	Teste de avaliação de cenários de interação com tarefas Questionário de satisfação de usuários	Estudo de acessibilidade e usabilidade em RDI de Universidades Federais brasileiras. Foi realizado teste de avaliação de cenários de interação com tarefas e questionário de satisfação de usuários
<i>Usability of digital repository software: A study of DSpace installation and configuration</i> ^[30]	Körber N, Suleman H.	2008	Questionário de satisfação de usuários	Estudo de usabilidade em RDI desenvolvido no DSpace. Mostra questionário para avaliar usabilidade da instalação, configuração do DSpace e criação de RDI
<i>Organizational Usability of Digital Libraries: Case Study of Legal Research in Civil and Criminal Courts</i> ^[31]	Elliott M, Kling R.	1997	Teste de avaliação de cenários de interação com tarefas	Estudo de caso de criação de um RD. Mostra as etapas da sua criação e a avaliação da usabilidade
<i>Intelligent e-learning repository system for sharing learning resources. In: Web and Open Access to Learning</i> ^[32]	Alanazi A, Abbod M, Ullah A.	2014	Teste de avaliação de cenários de interação com tarefas	Avaliação de RDI com exemplos de criação de cenários de interação com tarefas.
Physicians, guidelines, and cognitive tasks ^[33]	Bracha Y, Brottman GM, Carlson A.	2010	Teste de avaliação de cenários de interação com tarefas	Estudo de caso de avaliação de usabilidade de RD através de cenários de interação com tarefas para avaliação da usabilidade propõe a criação de um manual para auxiliar o usuário
<i>Use of Repositories of Digital Educational Resources. State of the Art Review</i> ^[34]	Rodes-Paragarino V, Gewerc-Barujel A, Llamas-Nistal M.	2016	Teste de avaliação de cenários de interação com tarefas	Evolução do uso de RDI e estudo de usabilidade

Referência [número da referência]	Autores	Ano	Avaliação da usabilidade	Resumo
<i>Health professional learner attitudes and use of digital learning resources</i> ^[35]	Maloney S, Chamberlain M, Morrison S, Kotsanas G, Keating JL, Ilic D	2016	Testes de inspeção por <i>checklist</i> com o grau de satisfação do usuário	Uma estratégia de avaliação em RDI baseada em teste de inspeção por <i>checklist</i>
<i>Comparison of software repositories for their usability in software process reconstruction</i> ^[36]	Janković M, Bajec M.	2015	Teste de inspeção por heurísticas Teste avaliação de usabilidade por carga de trabalho cognitivo	Avaliação do RD por teste de inspeção por heurísticas; e o teste avaliação de usabilidade por carga de trabalho cognitivo. Sugere um guia de como deve ser organizada a construção de um RD para evitar um processo de reconstrução
<i>Special Issue on Strategies to Improve the Usability of Learning Object Repositories</i> ^[37]	Casali A, Cechinel C, Ochoa X.	2016	Teste de avaliação de cenários de interação com tarefas	Estratégias para avaliação de usabilidade de RDI baseada em cenários de interação com tarefas
<i>Educational Repositories</i> ^[38]	Mahauad J, Maldonado J, Carvallo J, Zambrano J.	2016	Teste de avaliação de cenários de interação com tarefas	Evolução do uso de RDI e estudo de usabilidade
<i>Localizing OER in Afghanistan: Developing a Multilingual Digital Library for Afghan Teachers</i> ^[39]	Oates L, Hashimi J.	2016	Teste de avaliação de cenários de interação com tarefas	Estudo do desenvolvimento e avaliação da usabilidade em RD baseada em cenários de interação com tarefas

Discussão

Esta RIL mostrou que foram propostos alguns novos métodos de teste de inspeção de usabilidade^[2,25,28,36] e de avaliação de cenários de interação com tarefas.^[21,22,25,28,30,36]

Os novos métodos de teste de inspeção de usabilidade basearam-se em sua grande maioria em questionários que buscaram a opinião da satisfação do usuário em usar o RD ou RDI. Já os novos métodos de teste avaliação de usabilidade baseada em cenários de interação com tarefas considerou a carga de trabalho cognitivo utilizando a AET, que permite recolher informações específicas sobre uma dada tarefa, como por exemplo: objetivos, organização, procedimentos, *inputs* e resultados, terminologia, nível de produção, sistemas e ferramentas de trabalho,

manuals, comunicação, normas de qualidade, frequência, duração e aspectos críticos. Pode ser feita antes da concepção do sistema, (quando procura-se compreender como a tarefa é executada e recolher informações do novo sistema ou versão) durante (quando busca-se compreender ou definir como o usuário efetua a tarefa com o novo sistema) ou depois da concepção (quando se realizam estudos para compreender como os usuários fazem uso do sistema).^[1,11,10]

A observação do uso do sistema pelo usuário permite ao avaliador ter uma visão dos problemas sendo vivenciados pelos usuários e muitas vezes dos aspectos positivos experimentados durante o uso.

Alguns artigos adotaram outro método de coletar informações sobre como os usuários usam o sistema. Fizeram-no por meio de registros feitos durante o uso, com ou sem a presença concomitante do avaliador. Isto foi realizado por via de *logs*, que armazenam em um arquivo as ações executadas em um sistema, através da gravação da interação do usuário com o sistema, ou da gravação em vídeo da experiência do usuário.

Os artigos apontaram melhoria nas interfaces com a avaliação da usabilidade por *checklist* e por cenários de interação com tarefas. Alguns combinaram essas técnicas e outros apesar de não as realizarem sugerem essa combinação.

Considerações finais

A estratégia RIL nas bases de dados utilizadas e descritores usados mostrou que os testes aplicados melhoram a usabilidade dos RDs (28%) e apontam melhorias à interface gráfica com o uso do teste de inspeção por *checklist* (72%) e/ou por cenários de interação com tarefas (83%), isoladamente ou combinando as duas técnicas (33%) e, essa informação é relevante para a obtenção de uma interação efetiva entre o usuário e o material disponível em um RD a sua interface.

Mostrou também que entre os modelos, métodos e técnicas utilizados na avaliação de usabilidade de RDs foram de teste de inspeção por *checklist* (45%) ou teste de avaliação de cenários de interação com tarefas (50%).

A RIL permitiu verificar o estado da arte das técnicas de usabilidade aplicadas à ergonomia de RDs. Mostrou que foram propostos alguns novos métodos de teste de inspeção de usabilidade. Os novos métodos de teste de inspeção de usabilidade basearam-se, em sua grande maioria, em questionários que buscaram a opinião da satisfação do usuário em usar o repositório. Já os novos métodos de teste avaliação de usabilidade baseada em cenários de interação com tarefas considerou a carga de trabalho cognitivo, que permite recolher informações específicas sobre uma dada tarefa, como por exemplo: objetivos, organização, procedimentos, *inputs* e resultados, terminologia, nível de produção, sistemas e ferramentas de trabalho, manuals, comunicação, normas de qualidade, frequência, duração e aspectos críticos.

Alguns artigos adotaram outro método de coletar informações sobre como os usuários usam o sistema. Fizeram-no por meio de registros feitos durante o uso, com ou sem a presença concomitante do avaliador. Isto foi realizado através de

logs, que armazenam em um arquivo as ações executadas em um sistema, por via da gravação da interação do usuário com o sistema ou da gravação em vídeo da experiência do usuário.

Os artigos apontaram melhoria nas interfaces com a avaliação da usabilidade por *checklist* e por cenários de interação com tarefas. Alguns combinaram essas técnicas e outros, apesar de não as realizarem, sugerem essa combinação.

Referências

1. Santos W, Gamez L, Mancini F. Ergonomic Evaluation of the Portal of the Repository in the Health Area of Unifesp: Proposal of Specifications and Ergonomic Recommendations for Its Interface [Internet]. In: Antona M, Stephanidis C, organizadores. Universal Access in Human-Computer Interaction. Human and Technological Environments. Cham: Springer International Publishing; 2017 [citado 2017 out 8]. página 26–38. Available from: http://link.springer.com/10.1007/978-3-319-58700-4_3
2. Camargo LSA, Vidotti SBG. Uma estratégia de avaliação em repositórios digitais. In: XV Nacional de Bibliotecas Universitárias. São Paulo: 2008.
3. Prates R, Barbosa S. Avaliação de Interfaces de Usuário–Conceitos e Métodos [Internet]. In: Jornada de Atualização em Informática do Congresso da Sociedade Brasileira de Computação. Sociedade Brasileira de Computação; 2003 [citado 2017 set 25]. Available from: http://homepages.dcc.ufmg.br/~rprates/ge_vis/cap6_vfinal.pdf
4. Afonso AP, Lima JR, Cota MP. A heuristic evaluation of usability of Web interfaces. IEEE Inf Syst Technol CISTI 2012;1–6.
5. Arellano MA. Preservação de documentos digitais. Rev Ci Inf 2004;33(2):15–27.
6. Conselho Nacional de ArquivosNAL DE ARQUIVOS (CONARQ). Diretrizes para a implementação de repositórios digitais confiáveis de documentos arquivísticos. 2014;
7. Sayão LF. Repositórios digitais confiáveis para a preservação de periódicos eletrônicos científicos. PontodeAcesso 2011;4(3):68–94.
8. Soares SSKP. Elaboração de materiais científicos educacionais multimídia na área da saúde utilizando conceitos de design gráfico de interfaces, usabilidade e ergonomia [Internet]. 2015 [citado 2015 out 29]; Available from: <http://dspace.c3sl.ufpr.br/dspace/handle/1884/38174>
9. IEA. What is ergonomics [Internet]. 2015 [citado 2015 out 28]; Available from: <https://www.iea.cc/whats/>
10. Gamez L. A construção da coerência em cenários pedagógicos online: uma metodologia para apoiar a transformação de cursos presenciais que migram para a modalidade de educação à distância. 2004;
11. Cybis W, Betiol A, Faust R. Ergonomia e Usabilidade: conhecimentos, métodos e aplicações. 2º ed São Paulo: Novratec; 2010.

12. Scapin D, Bastien JMC. Ergonomic Criteria for Evaluating the Ergonomic Quality of Interactive Systems. In: Behaviour and Information Technology - BIT, 16 (4). 1997. página 220–3.
13. Nielsen J. Heuristic Evaluation. In: Usability inspection methods. Nova Iorque (EUA): NIELSEN, J; MACK, R.; 1994. página 25–62.
14. Ferreira SMSP. Repositório institucional em comunicação: o projeto REPOSCOM implementado junto à federação de bibliotecas digitais em Ciências da Comunicação 10.5007/1518-2924.2007 v12nesp1p77. Encontros Bibli Rev Eletrônica Bibl E Ciênc Informação 2007;12(1):77–94.
15. Santos HM, Flores D. Repositórios digitais confiáveis para documentos arquivísticos: ponderações sobre a preservação em longo prazo. Perspect Em Ciênc Informação 2015;20(2):198–218.
16. Rogers Y, Preece J, Sharp H. Design de interação: além da interação homem-computador. Porto Alegre, RS: Bookman; 2013.
17. Winckler M, Pimenta MS. Avaliação de usabilidade de sites web. Esc Informática SBC SUL ERI 2002 Ed Porto Alegre Soc Bras Comput SBC 2002;1:85–137.
18. Rodrigues E, Almeida M, Miranda Â, Guimarães AX, Castro D. RepositóriUM: criação e desenvolvimento do Repositório Institucional da Universidade do Minho. 2004 [citado 2015 out 26];Available from: <http://repositorium.sdum.uminho.pt/handle/1822/422>
19. LIMA I, SOUZA R, DIAS G. Interatividade e Usabilidade nas Bibliotecas Digitais no Processo Ensino-Aprendizagem. DataGramaZero 2012;13(3):A03.
20. Sales ESM, Bezerra EP, Pereira HB de B. Biblioteca digital SCOL: organização, gestão e difusão do conhecimento científico através de objetos de aprendizagem SCORM. Digital library SCOL: organization, management and diffusion of the scientific knowledge through learning objects SCORM [Internet] 2013 [citado 2015 nov 12];Available from: <http://repositorios.questoesemrede.uff.br/repositorios/handle/123456789/634>
21. Bohmerwald P. Uma proposta metodológica para avaliação de bibliotecas digitais: usabilidade e comportamento de busca por informação na Biblioteca Digital da PUC-Minas. Ciênc Informação Brasília 2005;34(1):95–105.
22. Veiga V, Pimenta DN, Machado R, Silva A, Silva CH da. Repositórios institucionais: avaliação da usabilidade na Fundação Oswaldo Cruz. XIV Encontro Nac Pesqui Em Ciênc Informação ENANCIB 2013 GT 11 Informação E Saúde [Internet] 2013 [citado 2016 mar 2];Available from: <http://www.arca.fiocruz.br/xmlui/handle/icict/8599>
23. Dias C. Usabilidade na web: criando portais mais acessíveis. 2º ed Rio de Janeiro - RJ: Alta Books; 2007.
24. Mendes KDS, Silveira RC de CP, Galvão CM. Integrative literature review: a research method to incorporate evidence in health care and nursing. Texto Amp Contexto - Enferm 2008;17(4):758–64.
25. Santos DB dos. Avaliação da usabilidade da interface do Lume Repositório Digital da UFRGS. 2013;

26. Oliveira ER. Avaliação ergonômica de interfaces da SciELO-Scientific Electronic Library Online. 2001;
27. Clayton N, Biddle R, Tempero E. A study of usability of Web-based software repositories [Internet]. In: Software Methods and Tools, 2000. SMT 2000. Proceedings. International Conference on. IEEE; 2000 [citado 2016 nov 2]. página 51–58. Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=890420
28. Hee Kim H, Ho Kim Y. Usability study of digital institutional repositories. Electron Libr 2008;26(6):863–881.
29. Marcelino C de S, Siebra S de A. Análise da interação de usuários com repositórios institucionais de instituições federais de ensino superior brasileiras. GESTÃOOrg - Rev Eletrônica Gest Organ - ISSN 1679-1827 [Internet] 2016 [citado 2018 abr 10];13(0). Available from: <https://periodicos.ufpe.br/revistas/gestaoorg/article/view/22123>
30. Körber N, Suleman H. Usability of digital repository software: A study of DSpace installation and configuration [Internet]. In: International Conference on Asian Digital Libraries. Springer; 2008 [citado 2016 out 4]. página 31–40. Available from: http://link.springer.com/chapter/10.1007/978-3-540-89533-6_4
31. Elliott M, Kling R. Organizational Usability of Digital Libraries: Case Study of Legal Research in Civil and Criminal Cours. JASIS 1997;48(11):1023–1035.
32. Alanazi A, Abbod M, Ullah A. Intelligent e-learning repository system for sharing learning resources [Internet]. In: Web and Open Access to Learning (ICWOAL), 2014 International Conference on. IEEE; 2014 [citado 2016 out 4]. página 1–5. Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7009209
33. Bracha Y, Brottman GM, Carlson A. Physicians, guidelines, and cognitive tasks. Eval Health Prof 2010;0163278710392981.
34. Rodes-Paragarino V, Gewerc-Barujel A, Llamas-Nistal M. Use of Repositories of Digital Educational Resources. State of the Art Review. IEEE Rev Iberoam Tecnol Aprendiz 2016;11(2):73–8.
35. Maloney S, Chamberlain M, Morrison S, Kotsanas G, Keating JL, Ilic D. Health professional learner attitudes and use of digital learning resources. J Med Internet Res 2013;15(1):e7.
36. Janković M, Bajec M. Comparison of software repositories for their usability in software process reconstruction [Internet]. In: 2015 IEEE 9th International Conference on Research Challenges in Information Science (RCIS). IEEE; 2015 [citado 2016 out 4]. página 298–308. Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7128890
37. Casali A, Cechinel C, Ochoa X. Special Issue on Strategies to Improve the Usability of Learning Object Repositories. Rev Iberoam Tecnol Aprendiz 2016;11(2):71–2.
38. Mahauad J, Maldonado J, Carvallo J, Zambrano J. Educational Repositories. IEEE Rev Iberoam Tecnol Aprendiz 2016;11(2):79–86.



39. Oates L, Hashimi J. Localizing OER in Afghanistan: Developing a Multilingual Digital Library for Afghan Teachers. Open Prax 2016;8(2):151–61.

Checklist de Inspeção Ergonômica de Repositórios Digitais Institucionais – CIERDI

Wilma Honório dos Santos¹, Prof. Dr. Felipe Mancini², Prof. Dr. Luciano Gamez³

¹ Mestre em Ciências. Universidade Federal de São Paulo (Unifesp). Email: wilma.santos@unifesp.br

² Gestão de Informática em Saúde. Universidade Federal de São Paulo (Unifesp)/ Universidade Aberta do Brasil (UAB). Email: fmancini@unifesp.br

³ Universidade Federal de São Paulo (Unifesp)/ Universidade Aberta do Brasil (UAB). Email: lucianogamez@unifesp.br

Resumo

Repositórios digitais podem apresentar lacunas como falhas de navegação, baixa usabilidade e acessibilidade. A partir desse contexto, torna-se necessária uma avaliação ergonômica da sua interface. Observou-se a inexistência de um *checklist* a ergonomia e a usabilidade de RDIs. Para suprir esta lacuna, este artigo propôs um *checklist* de inspeção ergonômica de repositórios digitais institucionais (CIERDI). Este trabalho teve como objetivo propor um *checklist* de especificações e recomendações ergonômicas para avaliar a usabilidade de repositórios digitais institucionais. Foi realizada uma revisão integrativa da literatura para avaliar o estado da arte das técnicas de usabilidade aplicadas à ergonomia de repositórios digitais institucionais e para subsidiar os métodos aplicados neste trabalho, o fichamento dos artigos incluídos na revisão integrativa da literatura para auxiliar na escolha do método de avaliação de usabilidade, seleção de um conjunto de requisitos para inspeção de usabilidade em repositórios digitais institucionais. Foi proposto um *checklist* de especificações e recomendações ergonômicas para avaliar a usabilidade de RDIs e identificar os pontos de adequação e inadequação de usabilidade e sugerir um conjunto de especificações e recomendações para sua interface.

Palavras-chave: repositórios digitais, repositórios digitais institucionais, ergonomia, usabilidade.

Introdução

Repositórios digitais (RDs) são ambientes informacionais para o armazenamento e gerenciamento de documentos digitais (DD) que permitem a organização e o acesso da produção científica e acadêmica, o gerenciamento de comunidades e coleções e facilitam a implementação das políticas e estratégias de preservação.^[1,2]

As instituições acadêmicas utilizam os repositórios digitais institucionais (RDIs) para dar suporte e gerenciar a informação científica, principalmente das atividades de pesquisa e ensino.

Com a ampla utilização dos computadores, o estudo da ergonomia foi ampliado para analisar a capacidade mental que possibilita às pessoas a produção, recuperação e compreensão das informações geradas pelas tecnologias digitais da informação e comunicação (TDIC) e fez surgir a ergonomia cognitiva.^[3,4]

Estudos realizados em RDIs mostram que eles podem apresentar lacunas, como falhas de navegação, baixa usabilidade, buscas limitadas, pouca divulgação do ambiente e pouca ou nenhuma utilização de serviços personalizáveis.^[2,5-10]

Neste contexto, estes devem ser avaliados no que tange à ergonomia e usabilidade para proporcionar uma interação efetiva entre o usuário, material disponível e a sua interface.^[3,11,12]

No decorrer deste estudo, observou-se a inexistência de um *checklist* focado nas especificações e recomendações ergonômicas para avaliar a ergonomia e a usabilidade de RDIs. Para suprir esta lacuna, este artigo propôs um *checklist* de inspeção ergonômica de repositórios digitais institucionais (CIERDI).

Objetivo

O objetivo deste trabalho é propor um *checklist* de especificações e recomendações ergonômicas para avaliar a usabilidade de RDIs.

Materiais e métodos

Para atender aos preceitos éticos em pesquisas conforme consta na Resolução nº 466/12 do Conselho Nacional de Saúde, este trabalho de pesquisa foi aprovado sob parecer de número 1.545.877 e CAAE 55570716.2.0000.5505, pelo Comitê de Ética em Pesquisa (CEP) do Setor de Ciências da Saúde da Universidade Federal de São Paulo (Unifesp).

Para alcançar o objetivo proposto neste trabalho, foram realizadas as seguintes etapas: revisão integrativa da literatura (RIL); escolha do método de avaliação de usabilidade e a seleção de um conjunto de requisitos para inspeção de usabilidade em RDIs baseados nas heurísticas e diretivas de Nielsen e colaboradores; classificação do conjunto de requisitos para inspeção de usabilidade de RDIs; a criação do CIERDI para inspeção de usabilidade de RDIs e a atribuição dos graus de severidade para cada requisito.

Revisão integrativa da literatura (RIL)

Para o levantamento do estado da arte das técnicas de usabilidade aplicadas à ergonomia de RDIs e para subsidiar os métodos aplicados neste trabalho, foi realizada uma revisão integrativa da literatura (RIL). Para nortear esta RIL, foram formuladas as seguintes perguntas de pesquisa: O uso de técnicas de usabilidade aponta melhorias

para a ergonomia de RDIs? Quais são os modelos, métodos e técnicas utilizados na avaliação de usabilidade de RDIs? Métodos de avaliação de usabilidade aprimoram a experiência do usuário quanto ao uso da interface?

Crítérios de inclusão

Esta RIL teve como critérios de inclusão: (1) artigos que continham técnicas e/ou métodos de avaliação ergonômica da usabilidade de RDs e RIDs; (2) artigos descritivos relacionando a avaliação da usabilidade com aplicações em RDs e/ou RDIs.

Crítérios de exclusão

Para aperfeiçoar a objetividade e a validade da pesquisa, excluíram-se artigos que se enquadraram em, pelo menos, um dos cinco critérios de exclusão: (1) artigos não escritos em português ou na língua inglesa; (2) artigos similares com resultados duplicados em diferentes bases de dados; (3) artigos incompletos (somente resumo, tutoriais, *whitepapers* ou *keynotes*); (4) capítulos de livros; (5) artigos que não continham alguma aplicação relacionada ao contexto de usabilidade em RDs e/ou RDIs.

String de busca

Para obtenção dos artigos do estudo, construiu-se incrementalmente uma *string* de busca:

#1 USABILIDADE OR USABILITY OR (ERGONOMIA COGNITIVA) OR (COGNITIVE ERGONOMICS) OR (AVALIAÇÃO ERGONÔMICA) OR (ERGONOMIC EVALUATION)

#2 (MÉTODOS DE AVALIAÇÃO) OR (EVALUATION METHODS) OR (EVALUATION) OR (METHOD*) OR (MÉTODO)

#3 (REPOSIT* DIGITA*) OR (REPOSIT* INSTITUTIONAL*) OR (REPOSIT* DIGIT* INSTITUTIONAL*) OR (REPOSIT*)

#4 #1 AND #2 AND #3

A *string* de busca foi baseada nos termos da questão de pesquisa: usabilidade, ergonomia cognitiva e RDs e RDIs. O processo de pesquisa consistiu em uma busca automática, via Portal de Periódicos CAPES/MEC¹ usando a Rede Privada Virtual (VPN) da Unifesp, na *Web of Science*, pois esta base de dados indexa mais de 12.000 revistas de impacto em todo o mundo, incluindo revistas como as da *Association for Computing Machinery (ACM) Digital Library* ou as do *Institute of Electrical and Electronics Engineers (IEEE)*. A pesquisa foi realizada no dia 26/09/2016.

¹ <http://www.periodicos.capes.gov.br/>

Avaliadores externos

Após a busca automática, os artigos resultantes foram: (1) pré-selecionados segundo o título e o resumo; (2) selecionados segundo os critérios de inclusão e exclusão.

Participaram desta RIL cinco pesquisadores que preencheram os seguintes critérios de inclusão: (1) graduado em nível superior; (2) cursando pós-graduação *stricto sensu*; (3) possuir conhecimento de Metodologia Científica e (4) apresentar disponibilidade para avaliação dos artigos.

Os artigos foram distribuídos aleatoriamente entre os pesquisadores que, após a leitura do material de apoio e dos artigos selecionados, responderam ao questionário de avaliação para compor a RIL.

Material de apoio

Foi criado um blog “*MITEADS - Métodos e Inovações Tecnológicas para Educação a Distância em Saúde*² com o seguinte material de apoio para transmitir aos pesquisadores o conhecimento sobre o assunto: conceitos de usabilidade, ergonomia, RDs e RDIs; tutorial *Zotero*³; relação completa das referências dos artigos a serem avaliados e o formulário de avaliação dos artigos selecionados criado no *Google Forms*⁴. Utilizou-se o gerenciador de referências *Zotero*⁵ de maneira compartilhada com os pesquisadores a fim de organizar e disponibilizar os artigos selecionados e auxiliar na composição das referências bibliográficas desta RIL, o gerenciador de planilhas *Microsoft Excel 2016* para a criação de tabelas dinâmicas e gráficos como ferramentas de apoio para melhorar a compreensão dos resultados.

Os artigos que satisfizeram os critérios de inclusão e com informação suficiente para identificar os modelos, métodos e técnicas utilizados para essa RIL e foram avaliados em forma de respostas a um questionário com questões pré-estabelecidas sobre o assunto, pertinentes à pergunta da pesquisa, algumas com alternativas afirmativas ou negativas ou a escala de *Likert*.

Fichamento dos artigos da revisão integrativa da literatura

Foram feitos a leitura e o fichamento dos 18 artigos incluídos na RIL para subsidiar a escolha do método de avaliação de usabilidade e seleção de um conjunto de requisitos para inspeção de usabilidade em RDIs.

²<http://miteds.blogspot.com.br/>

³ <https://www.zotero.org/support/pt/start>

⁴ <https://docs.google.com/forms/u/0/>

⁵ <https://www.zotero.org>

Criação do CIERDI

Após a seleção do conjunto de requisitos para inspeção de usabilidade em RDIs, foi proposto o *Checklist* de inspeção ergonômica de repositórios digitais institucionais (CIERDI) para o teste de inspeção de usabilidade, composto por um conjunto de requisitos, heurísticas, diretivas e graus de severidade baseados nas heurísticas de Nielsen e seus colaboradores, bem como suas diretivas para a Elaboração de Conteúdo, Produção de Páginas, Projeto Navegacional, Projeto Arquitetural e Projeto de Interface.^[13-16]

Para cada grau de severidade, foi atribuído um número de 0 a 4 que, de maneira crescente, representou o aumento no comprometimento da usabilidade no RDI, bem como na velocidade de solução da violação do requisito no CIERDI. Podemos dizer que o grau severidade igual a 0 indica que o problema encontrado não está relacionado à usabilidade. Para o grau de severidade 1, o problema refere-se à aparência da(s) página(s), o que significa que precisa ser corrigido somente se um tempo estiver disponível. Já o grau de severidade igual a 2 corresponde a um problema de usabilidade pequeno e possui baixa severidade para ser solucionado. Quando o grau de severidade corresponde a 3 significa que o problema de usabilidade é grande e é importante solucioná-lo, pois tem alta prioridade. Mas quando o grau de severidade corresponde a 4, temos um grave comprometimento de usabilidade e é imprescindível solucionar esse problema antes de o *site* entrar no ar ou o mais rápido possível, pois seu comprometimento é crítico, tanto para o RDI, como para o usuário, quando se refere aos resultados esperados e obtidos.

Resultados

Na etapa da RIL, foram utilizados 18 artigos que apresentaram os itens para responder à pergunta de pesquisa. O Quadro 1 mostra a compilação dos resultados da RIL.

Quadro 1: Compilação dos resultados da RIL

Compilação dos resultados da RIL	%
Apontam melhorias à interface gráfica usando cenários de interação com tarefas	83
Apontam melhorias à interface gráfica com o uso de teste de inspeção por <i>checklist</i>	72
Recomendam a combinação das duas técnicas	56
Proposta de um novo método de teste de avaliação de cenários de interação com tarefas	50
Avaliação da usabilidade por teste de cenários de interação com tarefas	50
Avaliação da usabilidade com o uso de teste de inspeção por <i>checklist</i>	45
Consideram o teste de avaliação de cenários de interação com tarefas mais eficiente	44
Proposta de um novo método de teste de inspeção de usabilidade por <i>checklist</i>	39
Combinam as duas técnicas	33
Consideram o teste de inspeção por <i>checklist</i> mais eficiente	33

Compilação dos resultados da RIL	%
Os testes aplicados melhoram a usabilidade do RD	28
Comparação das duas técnicas	28

A RIL permitiu identificar autores, livros e artigos que não contemplaram os critérios de inclusão, mas que se mostraram relevantes para a escolha de abordagens de avaliação de usabilidade e de um conjunto de requisitos para inspeção de usabilidade em RDI.^[14,15,17,18]

O fichamento de todos os artigos incluídos na RIL foi determinante para a eleição de abordagens, a avaliação de usabilidade e seleção de um conjunto de requisitos para inspeção de usabilidade em RDIs baseados nas heurísticas de Nielsen e seus colaboradores, bem como suas diretrizes para Elaboração de Conteúdo, Produção de Páginas, Projeto Navegacional, Projeto Arquitetural e Projeto de Interface. O CIERDI é mostrado a seguir no Quadro 2.

Quadro 2: CIERDI – Checklist de inspeção ergonômica de RDIs

Requisitos	Heurística	Diretiva	Grau de severidade	Resposta esperada
1 Exibe o nome do RDI e/ou logotipo na sua <i>homepage</i>	7	Projeto de Interface	1	Sim
2 O nome do RDI e/ou logotipo está localizado no canto superior esquerdo na sua <i>homepage</i>	7	Produção de Páginas	1	Sim
3 O seu logotipo está disponível em todas as páginas do RDI	7	Produção de Páginas	1	Sim
4 O seu logotipo é clicável e remete à <i>homepage</i> do RDI	7	Projeto Navegacional	1	Sim
5 O seu logotipo é clicável na <i>homepage</i> do RDI (<i>link</i> ativo para a <i>homepage</i> na <i>homepage</i>)	5	Projeto de Interface	2	Não
6 Existe uma <i>tagline</i> (frase explicativa) resumindo explicitamente o que o RDI faz	9	Projeto de Interface	1	Sim
7 O RDI apresenta o logotipo clicável da instituição de tamanho inferior ao seu e remete sua respectiva página	9	Elaboração do Conteúdo	1	Sim
8 As páginas do RDI mantêm os logotipos externos do lado direito	9	Produção de Páginas	2	Sim
9 As páginas do RDI mantêm os logotipos externos pequenos e o mais discreto possível em relação ao conteúdo central da <i>homepage</i> e ao logotipo do RDI	9	Produção de Páginas	2	Sim
10 O RDI possui <i>link</i> com a “Apresentação” ou “Sobre nós” que oferece aos usuários uma visão geral sobre o RDI	9	Projeto Navegacional	1	Sim
11 O RDI possui <i>link</i> com a “Política”	9	Projeto Navegacional	2	Sim
12 O RDI possui <i>link</i> com a “Política de privacidade”	9	Projeto Navegacional	2	Sim
13 O RDI possui <i>link</i> com a frase “Direitos autorais”, citando a legislação vigente	9	Projeto Navegacional	1	Sim
14 O RDI possui <i>link</i> com o certificado de segurança	9	Projeto Navegacional	2	Sim
15 O RDI possui a opção “entre em contato” ou “fale conosco” com todas	9	Projeto Navegacional	2	Sim

Requisitos	Heurística	Diretiva	Grau de severidade	Resposta esperada
as informações de contato do repositório				
16 A opção “entre em contato” ou “fale conosco” funciona	9	Projeto Navegacional	2	Sim
17 O RDI mostra a estrutura do <i>site</i> , possui um <i>link</i> “Mapa do <i>site</i> ”	1	Projeto Arquitetural	1	Sim
18 O RDI provê um mapa de <i>site</i> dinâmico que indica a página da qual foi acessada e tenha formas de destacar informações de interesse a populações de usuários específicas	1	Projeto Navegacional	2	Sim
19 O RDI possui a opção “Estatísticas”	1	Projeto Navegacional	2	Sim
20 O RDI possui o <i>link</i> “Ajuda”	10	Projeto Navegacional	3	Sim
21 O conteúdo do RDI agrupa informações do repositório, como “Sobre Nós”, “Apresentação”, “Política”, “Política de Privacidade” em uma única área reservada	7	Elaboração do Conteúdo	2	Sim
22 A <i>homepage</i> do RDI informa de modo claro a consistência das informações disponíveis	4	Elaboração do Conteúdo	2	Sim
23 O RDI deixa visíveis a estrutura e a forma como o RIUnifesp está organizado	4	Produção de Páginas	2	Sim
24 O RDI possui <i>links</i> permanentes	9	Produção de Páginas	1	Sim
25 O RDI funciona igualmente nos navegadores <i>Internet Explorer</i> , <i>Mozilla</i> , <i>Opera</i> e <i>Google</i>	1	Projeto Navegacional	3	Sim
26 O RDI usa a seta “voltar” do navegador ao invés de <i>link</i> próprio para essa finalidade	3	Projeto Navegacional	2	Sim
27 O RDI enfatiza as tarefas de mais alta prioridade para que os usuários tenham um ponto de partida definida na <i>homepage</i>	3	Projeto de Interface	2	Sim
28 O RDI não usa a palavra “ <i>website</i> ” para se referir a qualquer outro aspecto	9	Projeto de Interface	1	Sim
29 O RDI estrutura a sua <i>homepage</i> de modo diferente de todas as outras páginas existentes no <i>site</i>	9	Projeto de Interface	1	Sim
30 O RDI evita utilizar diversas caixas de entrada de texto na <i>homepage</i> , principalmente na parte superior da página em que as pessoas geralmente procuram o recurso de pesquisa	9	Projeto de Interface	1	Sim
31 O RDI usa raramente menus suspensos, principalmente se os itens neles contidos não forem autoexplicativos	9	Projeto de Interface	1	Sim
32 O RDI não apresenta <i>links</i> genéricos para suporte à comunidade de usuários, <i>chat</i> ou outros recursos para discussão	9	Projeto de Interface	1	Sim
33 O RDI não oferece uma entrada em “Livro de visitantes”, pois os faz parecer amadores	9	Projeto de Interface	1	Sim
34 O RDI inclui a palavra com o nome do idioma na própria língua (por exemplo, usar a palavra “ <i>English</i> ” como âncora para o <i>site</i> traduzido em inglês)	9	Projeto de Interface	1	Sim
35 O RDI evita o uso de bandeira para indicar um <i>site</i> traduzido, pois um país	9	Projeto de Interface	1	Sim



Requisitos	Heurística	Diretiva	Grau de severidade	Resposta esperada
36	9	Projeto de Interface	1	Sim
37	3	Projeto de Interface	2	Sim
38	9	Projeto de Interface	2	Sim
39	9	Projeto de Interface	2	Sim
40	9	Projeto de Interface	2	Sim
41	9	Projeto de Interface	2	Sim
42	9	Projeto de Interface	2	Sim
43	4	Projeto Arquitetural	1	Sim
44	4	Projeto Arquitetural	1	Sim
45	4	Projeto Arquitetural	1	Não
46	5	Projeto Arquitetural	1	Sim
47	9	Elaboração do Conteúdo	1	Sim
48	9	Elaboração do Conteúdo	1	Sim
49	9	Elaboração do Conteúdo	1	Sim
50	9	Elaboração do Conteúdo	1	Sim
51	9	Elaboração do Conteúdo	1	Sim
52	9	Elaboração do Conteúdo	1	Sim
53	9	Elaboração do Conteúdo	1	Sim
54	9	Elaboração do Conteúdo	1	Sim
55	1	Projeto Navegacional	1	Sim
56	2	Elaboração do Conteúdo	1	Sim
57	9	Elaboração do Conteúdo	1	Sim



Requisitos	Heurística	Diretiva	Grau de severidade	Resposta esperada
58 O conteúdo do RDI emprega letras maiúsculas e outros padrões de estilo com consistência	9	Elaboração do Conteúdo	1	Sim
59 O RDI não rotula uma área nitidamente definida da página se o conteúdo for suficientemente autoexplicativo	9	Elaboração do Conteúdo	1	Sim
60 O conteúdo do RDI evita itemização excessiva (listas com um único item)	9	Elaboração do Conteúdo	1	Sim
61 O conteúdo do RDI utiliza espaços não separáveis entre palavras que precisam permanecer juntas, para serem vistas e entendidas nas frases	9	Elaboração do Conteúdo	1	Sim
62 O conteúdo do RDI usa somente o discurso imperativo, como “Insira uma cidade ou CEP” nas tarefas obrigatórias, ou qualifica a declaração adequadamente	9	Elaboração do Conteúdo	1	Sim
63 O conteúdo do RDI explica o significado de abreviações, iniciais maiúsculas, acrônimos e segue-os imediatamente com as abreviações na primeira ocorrência	9	Elaboração do Conteúdo	1	Sim
64 O conteúdo do RDI evita pontos de exclamação	9	Elaboração do Conteúdo	1	Sim
65 O conteúdo do RDI emprega raramente todas as letras em maiúsculas e nunca como um estilo de formatação	9	Elaboração do Conteúdo	1	Sim
66 O conteúdo do RDI evita usar inadequadamente espaços e pontuação para dar ênfase	9	Elaboração do Conteúdo	1	Sim
67 O conteúdo do RDI apresenta títulos sucintos, mas descritivos, para transmitir o máximo de informações com um mínimo de palavras possível	9	Elaboração do Conteúdo	1	Sim
68 O RDI não disponibiliza para os usuários recursos para personalizar a aparência básica da interface da <i>homepage</i>	3	Elaboração do Conteúdo	1	Sim
69 O RDI não usa ícones que mostrem aos usuários gestos que sejam ofensivos em sua cultura	9	Elaboração do Conteúdo	1	Sim
70 O conteúdo do RDI não usa jogo de palavras visuais uma figura pode conter significados diferentes em culturas distintas	4	Elaboração do Conteúdo	1	Sim
71 O conteúdo do RDI não emprega metáforas fora do domínio da informação do <i>site</i>	4	Elaboração do Conteúdo	1	Sim
72 O conteúdo do RDI usa o nome do mês inteiro ou abreviações, mas não números	4	Elaboração do Conteúdo	1	Sim
73 As páginas do RDI produzem hipertextos produtivos	9	Produção de Páginas	1	Sim
74 As páginas do RDI respeitam a paragrafação	9	Produção de Páginas	1	Sim
75 Nas páginas do RDI, a quantidade de elementos gráficos é reduzida	9	Produção de Páginas	1	Sim
76 As páginas do RDI evitam gráficas de marca d'água (imagens de plano de fundo com texto sobreposto)	9	Produção de Páginas	1	Sim



Requisitos	Heurística	Diretiva	Grau de severidade	Resposta esperada
77 As páginas do RDI não usam a animação para o único propósito de chamar a atenção para um item na <i>homepage</i>	9	Produção de Páginas	1	Sim
78 As páginas do RDI jamais animam elementos críticos da página, como logotipo, <i>tagline</i> ou título principal	9	Produção de Páginas	1	Sim
79 As páginas do RDI limitam os estilos de fonte e outros atributos de formatação de texto, como tamanhos e cores	9	Produção de Páginas	1	Sim
80 As páginas do RDI usam texto com muito contraste e cores de plano de fundo para que os caracteres fiquem o mais legível possível	9	Produção de Páginas	1	Sim
81 As páginas do RDI usam fontes de tamanho suficiente para ler	9	Produção de Páginas	1	Sim
82 O plano de fundo das páginas do RDI é de cor branca	9	Produção de Páginas	1	Sim
83 O RDI possui barra de rolagem setas para baixo e para cima e um indicador de rolagem	9	Produção de Páginas	1	Sim
84 As páginas do RDI evitam a rolagem horizontal a 800x600	9	Produção de Páginas	1	Sim
85 Nas páginas do RDI os elementos mais críticos da página estão visíveis “acima da dobra” (na primeira tela de conteúdo, sem rolar), no tamanho de janela mais predominante	9	Produção de Páginas	1	Sim
86 As páginas do RDI usam um <i>layout</i> fluido para permitir o ajuste do tamanho da <i>homepage</i> a diversas resoluções de tela	9	Produção de Páginas	1	Sim
87 O RDI não inclui o nome de domínio de nível superior, como “.br”, no título da janela	9	Produção de Páginas	1	Sim
88 O RDI não inclui a palavra “Homepage” no título	9	Produção de Páginas	1	Sim
89 As páginas do RDI limitam os títulos das janelas a não mais do que sete ou oito palavras e a menos de 64 caracteres	9	Produção de Páginas	1	Sim
90 As páginas do RDI selecionam palavras com alto conteúdo informativo com âncoras de hipertexto	9	Produção de Páginas	1	Sim
91 O RDI deixa clara, desde o início, a abrangência do nicho que se deseja atender	9	Produção de Páginas	1	Sim
92 O RDI prove documentação de fácil uso	10	Produção de Páginas	1	Sim
93 O RDI evita <i>links</i> internos	9	Projeto Navegacional	1	Sim
94 O RDI facilita o acesso aos itens apresentados recentemente na <i>homepage</i> , como nas duas últimas semanas ou no mês anterior, fornecendo uma lista das últimas apresentações	9	Projeto Navegacional	1	Sim
95 O RDI não utiliza <i>links</i> profundos	9	Projeto Navegacional	1	Sim
96 No RDI os <i>links</i> estruturais são padrão em todo <i>site</i>	9	Projeto Navegacional	1	Sim
97 O RDI apoia a navegação controlada pelo usuário	9	Projeto Navegacional	1	Sim



Requisitos	Heurística	Diretiva	Grau de severidade	Resposta esperada
98 O RDI não usa instruções genéricas, como “clique aqui” como nome de um <i>link</i>	9	Projeto Navegacional	1	Sim
99 O RDI não usa <i>links</i> genéricos, como “Mais. . .”, no final de uma lista de itens	9	Projeto Navegacional	1	Sim
100 O RDI não usa a palavra “ <i>Links</i> ” para indicar <i>links</i> existentes na página	9	Projeto Navegacional	1	Sim
101 O RDI agrupa itens na área de navegação, de modo que os itens semelhantes fiquem próximos entre si	9	Projeto Navegacional	1	Sim
102 O RDI não inventa termos para as opções de navegação de categorias. As categorias são diferenciáveis entre si. Se os usuários não entenderem a terminologia inventada, não conseguirão distinguir as categorias	9	Projeto Navegacional	1	Sim
103 O RDI fornece um mecanismo de <i>feedback</i> , especifica o objetivo do <i>link</i> e outras informações pertinentes	1	Produção de Páginas	1	Sim
104 O RDI disponibiliza uma caixa de entrada na <i>homepage</i> para inserir consultas de pesquisa, em vez de oferecer apenas um <i>link</i> para uma página de pesquisa	9	Projeto Navegacional	1	Sim
105 A caixa de pesquisa do RDI é de cor branca	9	Projeto Navegacional	1	Sim
106 A caixa de pesquisa do RDI está posicionada em um lugar padrão em todas as páginas	9	Projeto Navegacional	1	Sim
107 O posicionamento da caixa de pesquisa do RDI é do lado esquerdo ou central, aproveitando a experiência do usuário em outros <i>sites</i> , e a área é limpa, conduzindo o olhar para esse campo	9	Projeto Navegacional	1	Sim
108 A caixa de busca não possui tamanho suficiente para os usuários verem e editarem consultas padrões no <i>site</i>	9	Projeto Navegacional	1	Sim
109 Na área de pesquisa com um título, usa a palavra “Busca” e não um símbolo da lupa à direita da caixa	9	Projeto Navegacional	1	Sim
110 O RDI não realiza verificações ortográficas tanto para os dados de entrada da busca como para termos dos documentos consultados	9	Projeto Navegacional	1	Sim
111 O RDI não oferece expansão de sinônimos para os dados da busca	9	Projeto Navegacional	1	Sim
112 O RDI fornece pesquisas simples na <i>homepage</i> , com um <i>link</i> para acessar a pesquisa avançada ou dicas de pesquisa	9	Projeto Navegacional	1	Sim
113 A caixa de busca a pesquisa avançada possui tamanho suficiente para os usuários verem e editarem consultas padrões no <i>site</i>	9	Projeto Navegacional	1	Sim
114 A caixa de busca a pesquisa avançada possui a cor branca	9	Projeto Navegacional	1	Sim
115 No RDI, o título selecionado para consulta ganha um destaque e um sublinhado	7	Projeto Navegacional	3	Sim
116 No RDI, os <i>links</i> ganham um destaque e um sublinhado	7	Projeto Navegacional	3	Sim

Requisitos	Heurística	Diretiva	Grau de severidade	Resposta esperada
117 Os textos das páginas do RDI são na cor preta	9	Projeto Navegacional	2	Sim
118 O RDI indica claramente quais <i>links</i> conduzem a informações de acompanhamento sobre cada exemplo e quais <i>links</i> direcionam para informações gerais sobre categoria como um todo	5	Projeto Navegacional	2	Sim
119 O RDI não utiliza sublinhado para textos que não sejam âncoras de hipertexto	9	Projeto Navegacional	2	Sim
120 O RDI permite <i>links</i> coloridos para indicar os estados visitados e não visitados	9	Projeto Navegacional	2	Sim
121 O RDI indica a presença de <i>links</i> com sublinhado e na cor azul	9	Projeto Navegacional	2	Sim
122 O RDI certifica-se de que o <i>link</i> indica exatamente o que acontecerá se for clicado (indicar se o <i>link</i> leva a outra página <i>web</i> , vincula um arquivo PDF a página, aciona um equipamento de áudio e vídeo ou um aplicativo de mensagens de <i>e-mail</i> etc.)	9	Projeto Navegacional	2	Sim
123 No RDI, a pesquisa na <i>homepage</i> deve pesquisar a site inteiro, por <i>default</i>	9	Projeto Navegacional	1	Sim
124 No RDI, quando se pesquisa uma palavra-chave que não existe, retorna-se essa informação	5	Projeto Navegacional	4	Sim
125 O RDI não oferece um recurso para “Pesquisar na <i>web</i> ”, na função de pesquisa do <i>site</i>	9	Projeto Navegacional	1	Sim
126 No RDI o resultado da busca: exibe a lista de ocorrência classificada com os melhores resultados situados na parte superior; elimina ocorrências repetidas de páginas iguais; exibe um pequeno trecho da página de destino, capaz de descrevê-la	9	Projeto Navegacional	1	Sim
127 O RDI possui um <i>link</i> “Mostrar o registro completo”	10	Projeto Navegacional	1	Sim
128 No RDI o <i>link</i> “Mostrar o registro completo” ganha um destaque e um sublinhado	7	Projeto Navegacional	3	Sim
129 No RDI existe um <i>link</i> com endereço permanente do arquivo	10	Projeto Navegacional	1	Sim
130 O <i>link</i> com endereço permanente do arquivo funciona	10	Projeto Navegacional	2	Sim
131 O RDI possui a opção “itens com mais <i>Downloads</i> ” após o resultado da busca	10	Projeto Navegacional	3	Sim
132 As páginas do RDI indicam entre parênteses o tamanho do arquivo e o formato, após o <i>link</i>	10	Produção de Páginas	1	Sim
133 No RDI, o arquivo do título selecionado abre na mesma janela	9	Projeto Navegacional	3	Sim
134 O RDI contempla os requisitos de usabilidade de documentos não- <i>Web</i> ,	4	Projeto Navegacional	3	Sim
135 O RDI disponibiliza o mesmo <i>link</i> em diversos locais da área de navegação	5	Projeto Navegacional	1	Sim
136 O RDI possui padronização de idioma e todas páginas	9	Produção de páginas	3	Sim
137 A navegação por título se encontra em ordem alfabética	9	Projeto Navegacional	1	Sim

Requisitos	Heurística	Diretiva	Grau de severidade	Resposta esperada
138 A navegação por palavra-chave se encontra em ordem alfabética	9	Projeto Navegacional	1	Sim
139 A navegação por autor se encontra em ordem alfabética	9	Projeto Navegacional	1	Sim
140 A navegação por data funciona	9	Projeto Navegacional	1	Sim

Discussão

Um alto grau de usabilidade em uma interface reflete que os usuários estão realizando tarefas com facilidade, rapidez e satisfação, que a interface proporciona facilidade de aprendizado, eficiência de uso, facilidade de memorização, baixa taxa de erros e satisfação subjetiva do usuário.^[25]

RIL

A RIL mostrou que foram propostos alguns novos métodos de teste de inspeção de usabilidade. Os novos métodos de teste de inspeção de usabilidade basearam-se, em sua grande maioria, em questionários que buscaram a opinião da satisfação do usuário em usar o RD. Já os novos métodos de teste avaliação de usabilidade baseada em cenários de interação com tarefas considerou a carga de trabalho cognitivo, que permite recolher informações específicas sobre uma dada tarefa, como por exemplo: objetivos, organização, procedimentos, *inputs* e resultados, terminologia, nível de produção, sistemas e ferramentas de trabalho, manuais, comunicação, normas de qualidade, frequência, duração e aspectos críticos.

Alguns artigos adotaram outro método de coletar informações sobre como os usuários usam o sistema. Fizeram-no por meio de registros feitos durante o uso, com ou sem a presença concomitante do avaliador. Isto foi realizado através de *logs*, que armazenam em um arquivo as ações executadas em um sistema, por via da gravação da interação do usuário com o sistema ou da gravação em vídeo da experiência do usuário.

Os artigos apontaram melhoria nas interfaces com a avaliação da usabilidade por *checklist* e por cenários de interação com tarefas. Alguns combinaram essas técnicas e outros, apesar de não as realizarem, sugerem essa combinação.

A RIL possibilitou o levantamento do estado da arte das técnicas de usabilidade aplicadas à ergonomia de RDIs e subsidiou os métodos aplicados.

Fichamento

O fichamento de todos os artigos incluídos na RIL orientou a escolha do método de inspeção de usabilidade escolhido e a seleção de um conjunto de requisitos para inspeção de usabilidade em RDIs e a criação do CIERDI.

CIERDI

A proposta do CIERDI veio preencher as lacunas existentes para a avaliação da usabilidade por teste de inspeção. Ele possibilitou a proposta de um *checklist* que contempla a inspeção da usabilidade de RDIs baseada nas heurísticas e diretivas de Nielsen e seus colaboradores. Foi possível atribuir a cada requisito um grau de severidade relacionado diretamente à urgência ou não da solução da sua violação e propiciou a avaliação, identificação dos pontos de adequação e inadequação e pode ser usado para avaliar RDIs.

Considerações finais

A proposta do CIERDI surgiu da RIL e do fichamento dos artigos incluídos nela. Nestas etapas, identificou-se uma lacuna de existência de um *checklist* com foco em teste de inspeção em RDIs. A RIL e o fichamento permitiram o direcionamento da escolha do método e seleção de um conjunto de requisitos para inspeção de usabilidade em RDIs base nas heurísticas e diretivas para Elaboração de Conteúdo, Produção de Páginas, Projeto Navegacional, Projeto Arquitetural e Projeto de Interface de Nielsen e seus colaboradores.

Dentre as principais contribuições apresentadas neste trabalho, pode-se elencar:

- A criação de *checklist* de inspeção ergonômica de repositórios digitais institucionais – CIERDI; a possibilidade de aplicação do CIERDI em RDIs; a busca da melhoria da usabilidade de outros RDIs; identificação dos pontos de adequação e inadequação de usabilidade em RDIs; e a diminuição da carga cognitiva do usuário da interface de RDIs.

Referências

1. Conselho Nacional de Arquivos (CONARQ). Diretrizes para a implementação de repositórios digitais confiáveis de documentos arquivísticos. 2014;
2. Sayão LF. Repositórios digitais confiáveis para a preservação de periódicos eletrônicos científicos. PontodeAcesso 2011;4(3):68–94.
3. Soares SSKP. Elaboração de materiais científicos educacionais multimídia na área da saúde utilizando conceitos de design gráfico de interfaces, usabilidade e ergonomia [Internet]. 2015 [citado 2015 out 29];Available from: <http://dspace.c3sl.ufpr.br/dspace/handle/1884/38174>
4. Cybis W, Betiol A, Faust R. Ergonomia e Usabilidade: conhecimentos, métodos e aplicações. 2o ed São Paulo: Novratec; 2010.
5. Ferreira SMSP. Repositório institucional em comunicação: o projeto REPOSCOM implementado junto à federação de bibliotecas digitais em Ciências da Comunicação 10.5007/1518-2924.2007 v12nesp1p77. Encontros Bibli Rev Eletrônica Bibl e Ciênc Informação 2007;12(1):77–94.
6. Camargo LSA, Vidotti SBG. Uma estratégia de avaliação em repositórios digitais. In: XV Nacional de Bibliotecas Universitárias. São Paulo: 2008.

7. Santos HM, Flores D. Repositórios digitais confiáveis para documentos arquivísticos: ponderações sobre a preservação em longo prazo. *Perspectivas em Ciência da Informação* 2015;20(2):198–218.
8. Sales ESM, Bezerra EP, Pereira HB de B. Biblioteca digital SCOL: organização, gestão e difusão do conhecimento científico através de objetos de aprendizagem SCORM. Digital library SCOL: organization, management and diffusion of the scientific knowledge through learning objects SCORM [Internet] 2013 [citado 2015 nov 12];Available from: <http://repositorios.questoesemrede.uff.br/repositorios/handle/123456789/634>
9. Bohmerwald P. Uma proposta metodológica para avaliação de bibliotecas digitais: usabilidade e comportamento de busca por informação na Biblioteca Digital da PUC-Minas. *Ciênc Informação Brasília* 2005;34(1):95–105.
10. Veiga V, Pimenta DN, Machado R, Silva A, Silva CH da. Repositórios institucionais: avaliação da usabilidade na Fundação Oswaldo Cruz. XIV Encontro Nac Pesqui Em Ciênc Informação ENANCIB 2013 GT 11 Informação E Saúde [Internet] 2013 [citado 2016 mar 2];Available from: <http://www.arca.fiocruz.br/xmlui/handle/icict/8599>
11. Afonso AP, Lima JR, Cota MP. A heuristic evaluation of usability of Web interfaces. *IEEE Inf Syst Technol CISTI* 2012;1–6.
12. Santos W, Gamez L, Mancini F. Ergonomic Evaluation of the Portal of the Repository in the Health Area of Unifesp: Proposal of Specifications and Ergonomic Recommendations for Its Interface [Internet]. In: Antona M, Stephanidis C, organizadores. *Universal Access in Human–Computer Interaction. Human and Technological Environments*. Cham: Springer International Publishing; 2017 [citado 2017 out 8]. página 26–38.Available from: http://link.springer.com/10.1007/978-3-319-58700-4_3
13. Nielsen J. Heuristic Evaluation. In: *Usability inspection methods*. Nova Iorque (EUA): NIELSEN, J; MACK, R.; 1994. página 25–62.
14. Nielsen J, Loranger H. *Usabilidade na Web : Projetando Websites com Qualidade*. Rio de Janeiro: Campus; 2006.
15. Nielsen J, Tahir M. *Homepage usabilidade: 50 websites desconstruídos*. Rio de Janeiro: Campus; 2012.
16. Nielsen J. Usability inspection methods [Internet]. In: *Conference companion on Human factors in computing systems*. ACM; 1994 [citado 2016 out 4]. página 413–414.Available from: <http://dl.acm.org/citation.cfm?id=260531>
17. Nielsen J. Heuristic Evaluation:How to Conduct a Heuristic Evaluation [Internet]. 1995 [citado 2018 abr 13];Available from: <https://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/>
18. Carvalho A, Anacleto J. *Usabilidade e seus critérios para a avaliação de sistemas computacionais*. São Carlos: Departamento de Computação da Universidade Federal de São Paulo; 2002.
19. Costa L, Ramalho F. A usabilidade nos estudos de uso da informação: em cena os usuários e sistemas interativos de informação. *Perspect Em Ciênc Informação* 2010;15(1):92–117.

Avaliação de confiabilidade do Repositório Institucional Arca

Michelle Lanzellote¹, Andrea Gonçalves do Nascimento², Claudete Fernandes de Queiroz³, Cláudio José Silva Ribeiro⁴

¹Farmanguinhos/Fiocruz, Rio de Janeiro, RJ. Email: michelle.lanzellote@hotmail.com

² ICICT/Fiocruz, Rio de Janeiro, RJ. Email: andrea.goncalves@icict.fiocruz.br

³ ICICT/Fiocruz, Rio de Janeiro, RJ. Email: claudete.queiroz@icict.fiocruz.br

⁴ UNIRIO, Rio de Janeiro, RJ. Email: claudio.ribeiro@unirio.br

Resumo

O trabalho discorre sobre um estudo realizado no Arca – Repositório Institucional da Fundação Oswaldo Cruz (Fiocruz) para a verificação de requisitos de confiabilidade do RI em questão, com base nas recomendações para repositórios digitais confiáveis. Utilizou-se o modelo de referência Open Archival Information System (OAIS) e os padrões de certificação TRAC, ISO 16363:2012 e CONARQ como base para a elaboração de um *checklist*, onde foram avaliados 27 critérios em três categorias: Infraestrutura organizacional, Gestão de objetos digitais e Infraestrutura tecnológica e de segurança, por meio da atribuição de uma escala de atendimento para cada critério. Como resultado do diagnóstico de confiabilidade, o RI Arca atende a 63% dos critérios de Infraestrutura organizacional, 27% dos critérios de Gestão de objetos digitais e 50% da categoria de Infraestrutura tecnológica e de segurança. Espera-se que esse estudo contribua para a implementação de melhorias no repositório Arca nos quesitos com baixo grau de avaliação, bem como o estabelecimento de padrões para quesitos implementados, porém não documentados. Espera-se ainda, que o diagnóstico realizado no RI Arca incentive outras instituições a iniciarem a autoavaliação de seus repositórios institucionais, visando uma futura auditoria e certificação.

Palavras-chave: Preservação digital; Repositórios digitais confiáveis; Certificação; Repositório institucional Arca.

Abstract

The work discusses a study carried out in the Arca - Institutional Repository of the Oswaldo Cruz Foundation (Fiocruz) to verify the trustability requirements of the IR, based on the recommendations for trustworthy digital repositories. The Open Archival Information System (OAIS) reference model and the TRAC, ISO 16363: 2012 and CONARQ certification standards were used as the basis to create a checklist with 27 criteria in three categories: Organizational Infrastructure, Management of Digital Objects and Technological and Security Infrastructure, by

assigning a service scale for each criterion. As a result of the diagnosis, the IR Arca meets 63% of the criteria for Organizational Infrastructure, 27% of the criteria for Management of Digital Objects and 50% for the Technological and Security Infrastructure category. We hope that this study will contribute to future improvements in the Arca repository within the low-grade evaluation criteria, as well as the establishment of standards for already implemented - but not documented - issues. We also expect that the diagnosis carried out at the IR Arca will encourage other institutions to begin the self-assessment of their institutional repositories, aiming at a future audit and certification.

Palavras-chave: Digital preservation; Trustworthy digital repositories; Certification; Institutional Repository Arca.

Introdução

Os repositórios institucionais se consolidaram na comunidade acadêmica nacional como uma importante ferramenta no ciclo de produção e circulação do conhecimento, cumprindo as funções de acesso e preservação da produção científica institucional. No entanto, um dos principais desafios enfrentados atualmente pelos repositórios está relacionado às práticas de preservação digital, desde a formulação de políticas institucionais até a implementação tecnológica de soluções que garantam a longevidade do conteúdo digital.

As instituições que abrigam repositórios institucionais assumem uma grande responsabilidade, pois devem ser capazes de garantir o acesso a seus acervos digitais de forma permanente, em um momento onde a preservação digital ainda é uma prática que deve ser aceita com base na *confiança* (CORRADO; SANDY, 2017), pois seus resultados (positivos ou não) serão conhecidos e comprovados somente no futuro. Os pesquisadores confiam que seus resultados de pesquisa estarão seguros no repositório institucional, e os gestores precisam garantir que o repositório atende a essa expectativa (HOUGHTON, 2015).

Segundo relatório da RLG/OCLC (2002), um repositório digital confiável é aquele cuja “missão é oferecer acesso confiável e de longo prazo aos recursos digitais por ele gerenciados para sua comunidade-alvo, agora e no futuro”. A forma encontrada pela comunidade acadêmica para estabelecer a confiança em um repositório digital é através do atendimento a determinados padrões e critérios que indicam evidência de confiabilidade, sendo isso devidamente documentado e, possivelmente, certificado.

Existem diversas iniciativas que buscam sistematizar evidências de que um repositório digital é confiável. Entre as mais reconhecidas estão a *Trustworthy Repositories Audit and Certification: Criteria and Checklist* (conhecida como TRAC), que serviu de base para a atual norma ISO 16363:2012 (*Audit and Certification of Trustworthy Digital Repositories*). No Brasil, temos ainda as diretrizes para a implementação de repositórios digitais confiáveis de documentos arquivísticos, do

Conselho Nacional de Arquivos (CONARQ). Neste trabalho, foi desenvolvida uma lista de requisitos comuns para a avaliação de confiabilidade de repositórios digitais com base no conjunto de critérios adotados por essas três normas, resultando em um *checklist*, que foi então aplicado ao Repositório Institucional Arca, da Fundação Oswaldo Cruz (Fiocruz).

Diversos autores (CORRADO; SANDY, 2017; HOUGHTON, 2015; MÁRDERO ARELLANO; OLIVEIRA, 2016) relatam que uma das vantagens da autoavaliação, assim como da auditoria de repositórios, é não somente poder transmitir confiança para seus usuários, mas também identificar os pontos fortes e as oportunidades de melhoria. Dessa forma, pretendemos com esse resultado fornecer subsídios para a constante melhoria do RI Arca, como primeiro passo rumo a um processo de certificação.

Considerando ainda que um levantamento feito por Arellano e Oliveira (2016) na literatura internacional mostra que a maioria dos repositórios institucionais está em seus primeiros estágios e apenas metade deles relata seguir algum processo de auditoria e certificação, esperamos contribuir também para que outras instituições iniciem seus processos de autoavaliação utilizando a ferramenta proposta.

Metodologia

Tendo o modelo de referência Open Archival Information System (OAIS) como fonte normativa para o *planejamento* de um repositório, sua *avaliação* se dá a partir de diferentes diretrizes, modelos e normas (BONAL ZAZO; DE LORENZO-CÁCERES, 2017), entre as quais se destacam os já mencionados critérios TRAC, a norma ISO 16363:2012 e, no Brasil, as diretrizes do CONARQ.

A partir da análise desses três documentos, foi feita uma lista de recomendações para repositórios digitais confiáveis, resultando em um *checklist* composto de 27 perguntas divididas em três categorias comuns aos três instrumentos: Infraestrutura organizacional, Gestão dos objetos digitais, e Infraestrutura tecnológica e de segurança.

Após a conclusão do *checklist*, foi feita a análise do repositório institucional Arca, da Fiocruz. O Arca tem como objetivo preservar a memória institucional, facilitar a pesquisa interdisciplinar, agilizar o processo da pesquisa, da descoberta e da inovação, garantindo o controle de autoria dos seus pesquisadores (FUNDAÇÃO, 2017). Desde o seu início, o número de registros inseridos no Arca aumentou consideravelmente, de cerca de 6.000 objetos em 2014, para mais de 26.000 em 2019.

Para analisar o Arca, foi preciso examinar em detalhe o ambiente virtual do repositório e a documentação disponível sobre sua organização, política, formas de uso e acesso. Todas as evidências utilizadas na avaliação foram obtidas *online*, em acesso aberto, como o Plano Operativo, o Termo de Uso, os Termos de Cessão, além de manuais e políticas que norteiam e regem o repositório. Também foram utilizados materiais sobre o Arca disponíveis em sua página web, como

apresentações em congressos e outros eventos, bem como uma entrevista complementar com o gestor do repositório, para validar dados e conclusões. Durante a aplicação do *checklist*, foi possível identificar informações disponíveis no repositório, porém não documentadas ou formalizadas institucionalmente e que, dessa forma, não puderam ser considerados como evidência. Com isso, destacamos a importância da documentação consistente do repositório, pois esta documentação é o principal insumo para qualquer processo de auditoria.

A confiabilidade do repositório foi mensurada através de uma escala do tipo Likert, que mede o nível de adesão a cada requisito, atribuindo um grau (de 0 a 4), conforme abaixo:

- Grau 0 – Não atende.
- Grau 1 – Atende a itens identificados como importantes na pergunta, mas não foi encontrado documento que comprove;
- Grau 2 – Atende a poucos itens identificados como importantes na pergunta;
- Grau 3 - Atende entre 50% e 90% dos itens identificados como importantes na pergunta;
- Grau 4 - Atende a todos os itens identificados como importantes na pergunta.

Os critérios para os quais não foram encontradas evidências receberam grau 0 ou 1, dependendo das informações não documentadas ou fornecidas pelo gestor.

Resultados e discussão

Apresentamos no Quadro 1 os critérios do *checklist* avaliado com os resultados obtidos, sua fundamentação e justificativa.

Quadro 1 – Checklist com aplicação, fundamentação, avaliação e justificativa dos quesitos

QUESITO	FUNDAMENTAÇÃO	AValiação	JUSTIFICATIVA
Infraestrutura Organizacional			
1. Estão claras e acessíveis a declaração de missão, documentação das decisões, desenvolvimento e ações do repositório bem como um histórico dessa documentação com todas as suas mudanças documentadas?	A declaração de missão está clara e acessível, pois contém o compromisso de longo prazo, as responsabilidades estão definidas e compartilhadas, assim como a documentação das decisões, desenvolvimento e ações do Arca estão estabelecidas na política, nas diretrizes e principalmente documentadas no plano operativo, que está disponível e acessível no repositório, bem como o histórico de atualização do plano operativo. (FUNDAÇÃO	Grau 4	Atende totalmente em todos os sentidos os critérios da questão.



	OSWALDO CRUZ, 2017; BRASIL, 2014)		
2. Existe uma boa governança e uma viabilidade organizacional com transparência de procedimentos, decisões, desenvolvimento e ações?	A partir do levantamento feito foi possível encontrar nas apresentações da coordenação do repositório em eventos e congressos, que estão disponíveis no ambiente virtual do Arca, dados estatísticos que comprovem a transparência de procedimentos, decisões, desenvolvimento e ações do repositório. Esses dados são enviados de 3 em 3 meses para o comitê de governança da política e do repositório, para acompanhamento de desempenho do repositório. Essa informação de periodicidade não encontrei em nenhum documento, mas é um ponto positivo a se destacar. (MARANHÃO et al., 2018)	Grau 4	Atende aos critérios da questão.
3. O repositório tem um número adequado de pessoal para suprir as funções e serviços de acordo com seu escopo? E possui programa de treinamento e desenvolvimento profissional contínuo?	De acordo com o plano operativo e as apresentações da coordenação do Arca, disponíveis no repositório. O Arca conta com uma equipe que supre as demandas do repositório com habilidades necessárias para operar nas áreas designadas e possui um programa de treinamento contínuo para os pesquisadores e bibliotecários. (FUNDAÇÃO OSWALDO CRUZ, 2017; MARANHÃO et al., 2018)	Grau 4	Atende aos critérios da questão.
4. A política existente para gestão do repositório é atual? Está instituída e disponível?	Faz parte de uma política instituída, atual e disponível. A política do Arca é a Política de Acesso Aberto ao Conhecimento da FIOCRUZ, foi instituída em 2014 e de lá para cá, não sofreu nenhuma atualização. Ela está disponível e acessível em todos os canais de	Grau 4	Atende totalmente os critérios de atualização e disponibilidade da questão.



	comunicação e informação da FIOCRUZ. (BRASIL, 2014)		
5. Em que medida as ações de auditoria e manutenção de políticas são implementadas e aplicadas regulamente?	O Arca não realiza ação de auditoria e não foi encontrada referência sobre esta questão nesta pesquisa.	Grau 0	Não atende aos critérios da questão.
6. O repositório tem a sua sustentabilidade financeira: planejamento financeiro de curto e longo prazo, transparente e claro quanto ao equilíbrio adequado de risco, benefício, investimento e despesa?	O Arca não tem um planejamento financeiro explícito, porque sua gestão financeira é no nível FIOCRUZ (como consta na política). Como o recurso é da instituição, dividido entre as suas unidades, cabe a cada unidade arcar com as despesas, principalmente no quesito, por exemplo, contratação de pessoal para trabalhar com o Arca. (BRASIL, 2014)	Grau 0	Não atende ao critério da questão e não foi possível encontrar informação documentada sobre a existência de nenhum planejamento financeiro no levantamento feito.
7. Os contratos, licenças e passivos estão claros, atualizados, acessíveis e disponíveis principalmente no que diz respeito à propriedade intelectual e a restrição de uso?	O Arca utiliza o termo de cessão, porque ele possui o caráter permanente, cede os direitos para a FIOCRUZ, cobrindo as questões de restrição de uso e a propriedade intelectual. Estão claros, atualizados de acordo com a demanda do repositório, acessíveis e disponíveis na intranet da FIOCRUZ. (FUNDAÇÃO OSWALDO CRUZ, 2017; BRASIL, 2014)	Grau 4	O termo de cessão atende aos critérios da questão com a função, objetivo e utilização dos documentos mencionados pela questão.
8. Existe algum tipo de contrato para suporte ao repositório?	O repositório utiliza o software DSpace, atualmente, na versão 4.7 caminhando para 6.3, mas sem contrato de suporte, pois não foi encontrada referência nesta pesquisa, que mostre que exista um contrato para o suporte.	Grau 0	Não atende ao critério da questão.
Gestão de Objetos Digitais			



9. Em que medida o repositório é compatível com padrões de metadados, como: Dublin Core, MODS, EAD, ANSI Z39.87, METS, PREMIS?	Segundo o plano de operação o Arca é compatível com o Dublin Core e o METS. (FUNDAÇÃO OSWALDO CRUZ, 2017)	Grau 2	Só adota dois padrões e pela importância deles em relação aos outros metadados.
10. O repositório tem controle sobre os objetos digitais que irá preservar? Especifica claramente a informação que precisa estar associada aos documentos no momento da submissão?	No plano operativo do Arca existe a descrição dos tipos de processos de depósitos, mas não detalha a questão da preservação, pois o repositório não tem um planejamento para preservação definido, descrito e publicado. (FUNDAÇÃO OSWALDO CRUZ, 2017)	Grau 3	Atende um dos dois quesitos da questão.
11. O repositório tem e usa identificadores únicos e visíveis de acordo com os padrões reconhecidos?	O repositório usa como identificador único e visível para todos os AIPs o <i>handle</i> e em alguns registros também o DOI. Esta informação foi identificada por observação, pois no levantamento não foi encontrada referência sobre esta questão. (FUNDAÇÃO OSWALDO CRUZ, 2017a)	Grau 1	Atende os itens identificados como importantes, mas não foi possível encontrar referência para documentar a informação observada.
12. Os procedimentos de adesão, os registros de processamento interno e de auditoria são transparentes?	No levantamento feito, não foi possível encontrar em nenhuma das referências detalhes do processamento interno e nem a confirmação se faz ou não auditoria.	Grau 0	Não atende aos critérios da questão.
13. Existem estratégias sólidas, atuais e documentadas de preservação, implementada e continuada, contendo os mecanismos de atualização dessas políticas identificados, com alertas sobre obsolescência iminente?	O Arca não possui um plano de preservação digital, pois este plano ainda está como um desafio ou perspectiva do repositório. Esta informação foi encontrada numa apresentação em slide da coordenação do Arca. (MARANHÃO et al., 2018)	Grau 0	Não atende aos critérios da questão.



14. O objeto digital está em PDF/A?	No plano operativo é recomendado a transformação do objeto digital para PDF/A antes de inserir no repositório, mas é visível que a maior parte dos objetos digitais do Arca não está em PDF/A. (FUNDAÇÃO OSWALDO CRUZ, 2017)	Grau 0	Não atende ao critério da questão.
15. Faz uso da migração, transformações, cópias, armazenamento distribuído e rastreamento do histórico de processamento que pode afetar a confiança na preservação?	De acordo com o plano operativo o Arca faz uso da migração, transformação, cópias e armazenamento distribuído. Inclusive tem o plano de <i>backup</i> descrito no plano operativo. Mas em relação ao rastreamento do histórico não foi localizado nenhum documento que comprove. (FUNDAÇÃO OSWALDO CRUZ, 2017)	Grau 3	Atende a maioria dos quesitos da questão.
16. Em que medida o repositório armazena, rastreia e usa metadados que correspondam às necessidades mínimas das comunidades designadas que suportem a sua funcionalidade básica?	De acordo com o plano operativo estão estabelecidos requisitos mínimos de metadados para cada tipologia do Arca, atendendo a funcionalidade básica de suas comunidades agregando ao objeto digital informação descritiva associada a ele. (FUNDAÇÃO OSWALDO CRUZ, 2017)	Grau 4	Atende aos critérios da questão.
17. Está documentado na Política o aspecto ou capacidade de acesso? E o que o usuário pode ou não fazer?	Está documentado na política os aspectos de acesso e no capítulo 3 e 4 da política estão descritos os direitos e deveres dos autores/usuários. O “Termo de uso” disponível na <i>homepage</i> do Arca, descreve o que o usuário/autor pode ou não fazer. (BRASIL, 2014)	Grau 4	Atende aos critérios da questão.
18. Há necessidade de diferentes políticas, para diferentes comunidades e diferentes tipos de	A política cobre a necessidade das diferentes comunidades. (BRASIL, 2014)	Grau 4	Como não há necessidade de criar diferentes políticas, os critérios da questão foram atendidos.



coleta? Se sim, já estão implementadas, instituídas e disponíveis?			
19. Há espaço para credenciar os usuários e algum mecanismo automatizado para identificar ameaças de segurança ou falhas no sistema de gerenciamento de acesso?	Há um espaço para cadastramento de usuário na página do repositório, mas não foi possível encontrar nenhum documento que prove a existência de um mecanismo que identifique ameaça de segurança ou falha no sistema, pois como o 'sistema' fica na 'sala cofre' da instituição e sob a guarda do setor da TI, qualquer irregularidade é comunicada à coordenação do Arca, que toma as providências cabíveis. (FUNDAÇÃO OSWALDO CRUZ, 2017a)	Grau 1	Só atende um dos critérios da questão, porque não foi possível encontrar informação documentada sobre procedimentos ligados a ameaças de segurança no levantamento feito.
Infraestrutura Tecnológica e de Segurança			
20. A Infraestrutura do sistema é segura e confiável, compreende onde estão os riscos?	No levantamento feito, não foi possível encontrar em nenhum dos documentos a confirmação que o sistema é seguro, confiável e compreende onde estão os riscos.	Grau 0	Não atende aos critérios da questão.
21. O Sistema é capaz de identificar o número de cópias, e suas sincronizações, de todos os objetos digitais armazenados e a sua localização, escrita sem ambiguidade?	No levantamento feito, não foi encontrada referência sobre a confirmação se o sistema identifica ou não.	Grau 0	Não atende aos critérios da questão.
22. Existe algum mecanismo eficaz para detectar corrupção ou perda de bits?	De acordo com uma apresentação sobre curadoria do repositório em um dos eventos da instituição é possível detectar a corrupção e a perda de bits através do trabalho de curadoria dos dados e pelo processo de migração. Esta apresentação está disponível no Arca na	Grau 4	Atende a todos os critérios da questão.



	comunidade do ICICT. (MARANHÃO et al., 2017)		
23. Demonstra a adequação dos processos <i>hardware</i> e <i>software</i> para seus sistemas de <i>backup</i> ?	De acordo com o plano operativo existem 2 tipos de <i>backup</i> que rodam em dias e horários diferentes: o <i>Full</i> e Incremental. (FUNDAÇÃO OSWALDO CRUZ, 2017)	Grau 4	Atende aos critérios da questão.
24. Existem estratégias e padrões relevantes para suas comunidades designadas e suas tecnologias digitais?	A estratégia de gestão e o seu funcionamento é igual para todas as comunidades e estão descritos no plano operativo. E na página inicial de cada comunidade tem a informação básica sobre ela. (FUNDAÇÃO OSWALDO CRUZ, 2017; FUNDAÇÃO OSWALDO CRUZ, 2017a)	Grau 4	Atende a todos os critérios da questão.
25. O sistema do repositório possui a tecnologia apropriada para garantir a Interoperabilidade entre as ferramentas de gestão, preservação e acesso?	O Arca possui a interoperabilidade entre os sistemas que o OAI-PMH permite. De acordo com o plano operativo, ele também interopera com os sistemas das bibliotecas, direto dos catálogos das bibliotecas para o Arca-produção. (FUNDAÇÃO OSWALDO CRUZ, 2017)	Grau 4	A interoperabilidade é automática.
26. O repositório tem requisitos de segurança estabelecidos como um plano escrito de preparo e recuperação de desastres, incêndio, inundações, comprometimento do sistema etc?	O Arca não possui um plano de segurança. Assim como na parte financeira o plano de segurança é no nível FIOCRUZ, como consta na política, abrangendo todos os sistemas, que ficam na 'sala cofre' da instituição e sob a guarda e responsabilidade do setor da TI. (BRASIL, 2014)	Grau 0	Não atende aos critérios da questão. No levantamento da pesquisa, não foi encontrada documentação sobre procedimentos ligados ao plano de segurança para o Arca.
27. Esses requisitos incluem pessoas que tem responsabilidades pelas ações, um <i>backup</i> fora do local da informação preservada e uma cópia fora do local do plano de preservação?	O Arca não possui um plano de segurança próprio, mas faz parte dos sistemas que ficam na 'sala cofre' da instituição e sob a guarda e responsabilidade do setor da TI. Embora o plano operativo do Arca estabeleça a rotina de <i>backup</i> , não é especificada sua localização. (BRASIL, 2014)	Grau 0	Não atende aos critérios da questão. No levantamento da pesquisa, não foi encontrada documentação sobre procedimentos ligados ao plano de segurança para o Arca.

Fonte: Dados da pesquisa (2018)

O resumo da avaliação segundo as categorias do *checklist* é apresentado na Tabela abaixo:

Tabela 1 – Resultado da avaliação do repositório institucional Arca por quantitativo de critérios em cada grau, de acordo com as categorias do *checklist*

Categorias	Grau 0	Grau 1	Grau 2	Grau 3	Grau 4
Infraestrutura Organizacional	3 (37%)	-	-	-	5 (63%)
Gestão de Objetos Digitais	3 (27%)	2 (18%)	1 (9%)	2 (18%)	3 (27%)
Infraestrutura Tecnológica e de Segurança	4 (50%)	-	-	-	4 (50%)

Fonte: Elaborado pelos autores

A categoria onde o repositório Arca obteve o maior número de critérios com grau 4 (63%) foi em Infraestrutura Organizacional. Esse resultado está em consonância com a observação de Arellano e Oliveira (2016) de que “os aspectos organizacionais da preservação digital são mais desafiadores que os assuntos técnicos, por isso muitos membros da comunidade têm se centrado nesses aspectos”.

A categoria Gestão de Objetos Digitais foi onde houve melhor distribuição dos requisitos, embora ainda apresente lacunas, com 45% dos critérios avaliados com grau 0 ou 1. A categoria Infraestrutura Tecnológica e Segurança teve resultado polarizado, porém recebeu o maior número de itens com o menor grau de avaliação (50%).

Em alguns casos, a avaliação do repositório foi prejudicada por falta de evidência documental que comprovasse o atendimento ao critério. Por exemplo, foi notada a falta de um plano de preservação digital, que é um dos principais instrumentos para garantir a confiança e a segurança de qualquer repositório digital, e ainda está em fase de desenvolvimento pela equipe do RI Arca.

Vale ressaltar que nem todos os critérios devem ter peso equivalente na avaliação geral de um repositório, sendo alguns muito mais importantes para a garantia da confiabilidade que outros (HOUGHTON, 2015).

Conclusão

A análise dos critérios de confiabilidade definidos pela TRAC, ISO 16363 e CONARQ, bem como os resultados apresentados, representam uma pequena parcela dos requisitos para que um repositório digital possa garantir sua confiabilidade e segurança, por isso há necessidade de desenvolver estudos mais aprofundados sobre o tema, sobretudo adaptados ao contexto nacional.

É importante lembrar que uma autoavaliação não é uma auditoria (HOUGHTON, 2015), porém sua realização é parte importante do processo de atualização e fortalecimento dos repositórios, incluindo sua estratégia de preservação digital, além de ser um pré-requisito desejável em preparação para uma futura auditoria externa para certificação de confiabilidade.

Em relação aos resultados da avaliação do repositório institucional Arca, recomendamos incorporar a seu planejamento futuro a implementação de melhorias nas áreas que receberam baixo grau de avaliação, bem como documentar institucionalmente as boas práticas e procedimentos que já estão implementados, porém não formalizados. Recomendamos ainda efetuar uma avaliação mais criteriosa, utilizando padrões mais próximos dos programas de certificação, e estabelecer um cronograma de reavaliação periódica do repositório.

Com esse exemplo específico, esperamos também contribuir para que outras instituições realizem a autoavaliação de seus repositórios institucionais, permitindo identificar pontos fracos e fortes e definir planos de desenvolvimento, com vistas a futura auditoria e certificação.

Referências

BONAL ZAZO, J. L.; DE LORENZO - CÁCERES, M. DEL P. O. Criterios de certificación y auditoría de repositorios digitales seguros en archivos. In: **Da produção à preservação informacional: desafios e oportunidades**. [s.l.] Publicações do Cidehus, 2017. p. 529–550. Disponível em: <https://books.openedition.org/cidehus/2835>. Acesso em: 17 abr. 2019.

BRASIL. FUNDAÇÃO OSWALDO CRUZ. **Portaria 329/2014-PR**. Institui a Política de Acesso Aberto ao Conhecimento, visando garantir à sociedade o acesso gratuito, público e aberto ao conteúdo integral de toda obra intelectual produzida pela Fiocruz. Rio de Janeiro, março de 2014. Disponível em: https://portal.fiocruz.br/sites/portal.fiocruz.br/files/documentos/portaria_politica_de_acesso_aberto_ao_conhecimento_na_fiocruz.pdf. Acesso em: 30 jun. 2017.

CORRADO, E. M.; SANDY, H. M. **Digital preservation for libraries, archives and museums**. 2. ed. Lanham, Mariland: Rowman & Littlefield, 2017.

FUNDAÇÃO OSWALDO CRUZ. Plano operativo: Arca repositório institucional: versão 2 outubro/2017. Rio de Janeiro, 2017. 23 p. Disponível em: https://www.arca.fiocruz.br/bitstream/icict/23055/10/Plano_Operativo_Arca_2017_v2.pdf. Acesso em: 17 jul. 2017.

HOUGHTON, B. Trustworthiness: Self-assessment of an Institutional Repository against ISO 16363-2012. **D-Lib Magazine**, v. 21, n. 3/4, p. 1–5, 2015. Disponível em: <http://www.dlib.org/dlib/march15/houghton/03houghton.html>. Acesso em: 27 mar. 2019.

MARANHÃO, Ana Maria Neves; ARAUJO, Luciana Danielli de; QUEIROZ, Claudete Fernandes de. Arca - Repositório Institucional da Fiocruz: 2011-2018. In: ENCONTRO CAPES DE CIÊNCIA ABERTA - REPOSITÓRIOS DIGITAIS, 1., 2018, Brasília, DF. **Anais...** Brasília, DF: CAPES, 2018. 35 p. Disponível em: <https://www.arca.fiocruz.br/handle/icict/28964>. Acesso em: 05 jul. 2017.

MÁRDERO ARELLANO, M. Á.; OLIVEIRA, A. F. DE. Gestão de repositórios de preservação digital. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, v. 14, n. 3, p. 465, 2016. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8646346>. Acesso em: 13 mar. 2019.

RLG/OCLC. **Trusted digital repositories**: attributes and responsibilities. Mountain View, CA.: Research Library Group/OCLC, 2002. Disponível em: <https://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>. Acesso em: 20 abr. 2019.

Rumbo a la autoevaluación: la elaboración de un Manual de Procedimientos de Digitalización

Carlos Javier Nusch¹, Marisa Raquel De Giusti², Esteban Cristian Fernández³, Ariel Lira⁴

¹ Proyecto de Enlace de Bibliotecas y Servicio de Difusión de la Creación Intelectual, Universidad Nacional de La Plata, Argentina y Centro de Servicios en Gestión de Información, Comisión de Investigaciones Científicas de la Provincia de Buenos Aires, Argentina. Correo: carlosnusch@prebi.unlp.edu.ar

² Proyecto de Enlace de Bibliotecas y Servicio de Difusión de la Creación Intelectual, Universidad Nacional de La Plata, Argentina y Centro de Servicios en Gestión de Información, Comisión de Investigaciones Científicas de la Provincia de Buenos Aires, Argentina. Correo: marisa.degiusti@sedici.unlp.edu.ar

³ Proyecto de Enlace de Bibliotecas y Servicio de Difusión de la Creación Intelectual, Universidad Nacional de La Plata, Argentina. Correo: estebancfernandez@prebi.unlp.edu.ar

⁴ Proyecto de Enlace de Bibliotecas y Servicio de Difusión de la Creación Intelectual, Universidad Nacional de La Plata, Argentina y Centro de Servicios en Gestión de Información, Comisión de Investigaciones Científicas de la Provincia de Buenos Aires, Argentina. Correo: alira@sedici.unlp.edu.ar

Resumen

Introducción: El planeamiento de la preservación dentro de un repositorio es un elemento clave de la autoevaluación y de la evaluación externa y exige confirmar lo realizado a través de documentos que deben ser creados y revisados, deben tener además, delimitadas las responsabilidades y los responsables como parte de un marco sistemático de buenas prácticas de gestión documental. **Materiales y metodología:** El siguiente trabajo describe de manera abreviada pero en toda su extensión, un Manual de Procedimientos de Digitalización y busca dar cuenta de las actividades que se realizan en un repositorio cuando se recibe una nueva obra para su digitalización y pretende ser un insumo útil para un proceso de autoevaluación e incluso de auditoría externa. Las obras recibidas pueden ser materiales relativamente nuevos pero sin una contraparte digital o materiales muy antiguos, con diferentes tipos de encuadernación, distinto estado de conservación, e incluso obras no textuales, como planos de obras arquitectónicas. **Resultados y discusión:** El desafío del Manual está en describir todo el flujo de trabajo: 1. El ingreso del material, la prioridad asignada, el depósito de la obra física y el seguimiento de los procesos de digitalización; 2. Los detalles del proceso de digitalización en sí mismo: software, hardware, cámara y escáner; 3. Las transformaciones a las que se someten los archivos tras la digitalización, las distintas copias generadas y los diferentes formatos de archivos utilizados; 4. El modo de trabajo en relación al guardado de los archivos.

Abstract

Introduction: The repository preservation planning is a key subject in auto-evaluation and external evaluation and requires to check the conducted tasks through documents which must be created, revised and inspected and also, must include delimited accountabilities and the responsible personnel as part of a systematic frame of good practices in document management. **Materials and methods:** This work describes briefly but completely a Digitalization Procedure Handbook and pursues to account the repository performed activities when a new document is received for a digitalization process and pretends to be a useful asset to an auto evaluation process or even an external repository auditing. The receipt documents are usually relatively new materials which has no digital version or even old documents with several kinds of binding, varied state of material preservation, and all the more, non textual works such as architectural plan. **Results and discussion:** The Handbook's challenge consists in the description of the digitalization workflow: 1. The documents income, the assigned priority, the physical document storage, and the digitalization processes tracing; 2. The digitalization process details itselfs: software, hardware, cámaras and scanners; 3. The different necessary transformations to which the archives are submitted after the image capture, the several digital copies and different archive formats employed; 4. The work method in order to archive storage.

Palabras clave: Digitalización, Preservación Digital, Procedimientos, Autoevaluación

Introducción

Los repositorios institucionales por su propia constitución deben asegurar la preservación y el acceso a largo plazo, en este sentido, el estándar ISO 14721 es el centro de atención de quien necesita cumplir con tales compromisos. El Modelo de Referencia OAIS es, además un eje central para la Norma ISO 16363 de Auditoría y certificación de repositorios de confianza que es la guía para evaluar si un repositorio es confiable.

Si bien la mayor parte de los contenidos con los que trabaja un repositorio institucional como SEDICI o CIC Digital ya han nacido en formato digital, existen otros materiales que requieren un proceso de digitalización de calidad y siguiendo un procedimiento de trabajo normalizado. La Universidad Nacional de La Plata cuenta con 17 unidades académicas con sus bibliotecas, la biblioteca pública y una red de museos que aloja una ingente cantidad de obras físicas, en papel e incluso objetos, de gran valor patrimonial y pasibles de ser digitalizados.

En el programa Memoria del Mundo (Edmondson, 2002) de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) se remarca la importancia del patrimonio documental como memoria colectiva y documentada de los pueblos, y la necesidad de que éste sea preservado y protegido. La sola preservación del material en formato digital no es suficiente, sino también su accesibilidad equitativa y su distribución, como la de sus productos derivados sea amplia y sirva para generar conciencia acerca de la importancia del patrimonio documental (Foster, Russell, Lyall, & Marshall, 1995, p. 5). Del mismo modo que realizar copias de seguridad de los archivos no es hacer preservación digital (Pérez Pineda, 2017, p. 3) el mero hecho de escanear un documento no asegura la correcta preservación y difusión a largo plazo. El proceso de digitalización

del patrimonio documental de una institución como la Universidad Nacional de La Plata implica procedimientos complejos que involucran a personal de diferentes áreas y unidades académicas trabajando en pos del cuidado de materiales delicados cuyo daño o pérdida serían enormemente lamentables.

En el ámbito de los archivos y bibliotecas, la conservación implica todas aquellas medidas destinadas a proteger adecuadamente los documentos, con el fin de prolongar su utilización en condiciones óptimas durante el mayor tiempo posible (Pérez Pineda, 2017, p. 7). En el caso del patrimonio documental de una universidad o institución educativa o de investigación, el proceso de digitalización debe contemplar la conservación del soporte de la obra debido a su importancia como objeto y documento histórico (un libro antiguo, un incunable, etcétera) además de lo concerniente a la preservación del documento digital.

Por estas razones se dio inicio a la elaboración de un Manual de Procedimientos de Digitalización que abarcara toda la documentación resultante de diferentes tareas realizadas por el personal del repositorio a los largo de los últimos años en lo concerniente a la digitalización de materiales y su preservación (De Giusti, Peloché, Nusch, & Urbina, 2016; Nusch et al., 2018a, 2018b). Dicho manual se encuentra actualmente en proceso de redacción y que pretende lograr la estandarización de las prácticas llevadas a cabo por el equipo de Digitalización tanto de SEDICI como de CIC Digital.

Hardware y Software

El SEDICI contó siempre con escáneres estándares a los que se les daba un uso esporádico. Primeramente se trabajaba con un HP modelo 5590 que luego fue reemplazado por el modelo HP N8400. Actualmente, el escáner más elemental, utilizado para digitalizar documentos es el HP 7500. En 2017 se sumó un escáner de alimentación automática (ADF) modelo Fujitsu Fi 6170. Ambos escáneres permitieron la captura de mayor cantidad de páginas con una buena calidad. En 2019 se sumó un escáner de grandes dimensiones modelo Contex IQ Quattro para el escaneo de planos y documentos de gran tamaño.

A su vez, el personal venía trabajando en el armado de diferentes escáneres de bajo costo con el objetivo de escanear materiales cuya encuadernación no permitiera el uso de escáneres corrientes. En Agosto de 2012 se firmó un Acuerdo entre Wikimedia Argentina y SEDICI con el objeto de realizar un programa de asistencia y colaboración recíprocas para la ejecución de tareas orientadas a la transferencia de tecnologías de digitalización, capacitación en digitalización y disponibilización virtual de contenidos. En el marco de dicho acuerdo se recibieron capacitaciones para el manejo del escáner DAL (Derecho A Leer), así como al proceso de trabajo de digitalización de obras y contenidos y temas referidos a la compatibilidad de la propiedad intelectual de los contenidos con las licencias Wikimedia Commons. El escáner DAL fue desarrollado por el grupo Derecho a Leer para la Facultad de Filosofía y Letras en julio de 2011. Inspirados en la experiencia de Do It Yourself desarrollaron un escáner para libros de forma artesanal y muy económica, mediante madera y cámaras digitales estándares para ofrecer soluciones de digitalización de acervos bibliográficos a bajo costo. El primer escáner DAL armado en 2012 para SEDICI se basó en el Model 1 de DIY y en el primer scanner DAL de la Universidad de Buenos Aires. En 2017 fue armado segundo tipo basado en el modelo Archivista 2014 de DIY que presentaba varias ventajas: mejora en el soporte del libro, mucho más firme y con la cama móvil, mejora en el soporte de las cámaras, menos flexible que el modelo anterior pero más estable, la

iluminación con lámparas LED y mejor fidelidad en la captura de colores, cámaras reflex de mayor calidad con lentes intercambiables.

A pesar de contar con un nuevo escáner Archivista 2014, el estado de la encuadernación de algunos materiales que debían escanearse presentaban especiales requerimientos de manipulación y tal situación hizo necesario el diseño ad hoc de un tercer escáner. Uno de los principales problemas era las dimensiones de la base donde se sitúan los libros y la distancia al cristal, que impedían la libertad de movimientos necesaria para un material en estas condiciones. De usar el modelo Archivista 2014 giro de las páginas durante el proceso de escaneo hubiera provocado el deterioro por su delicado estado y el peso del libro habría podido dañar la encuadernación textil del lomo.

Teniendo en cuenta las precauciones necesarias para el manejo del material se optó por un sistema de escaneo rediseñado a partir del Model 1 de DIY, con una cámara cenital apuntando hacia el libro, junto con dos luces LED dicroicas, las luces led tienen varias ventajas con respecto a las luces halógenas e incandescentes utilizadas en el pasado. Consumen hasta un 90 % menos, no generan calor considerable y tienen un Índice de reproducción cromática [CRI](#)¹ alto. (FADGI, 2016: 17). Uno de los brazos del atril del DAL se adaptó para poder sostener el libro siempre abierto a 150° aproximadamente, y balancear el peso de ambos lados de modo que no forzara la encuadernación.

Se dotó al Scanner de cámaras Nikon D5600, este tipo de cámara reflex permite a diferencia de las cámaras del primer Dal Canon sx130, el intercambio de lente. En la actualidad contamos con un lente del tipo zoom 18-55mm, dicho objetivo es muy versátil y se adaptó bien a esta labor, aunque un lente de focal fija de mayor luminosidad puede dar aún mejores resultados. También cuenta con formato de sensor DX de 24,2 Megapíxeles que permite obtener documentos escaneados a 400dpi y respetar los estándares de digitalización internacional como los de Library of Congress.

Es importante notar que la distancia focal en este lente es utilizada sobre los 35mm que es donde ya no presenta problemas de distorsión de la imagen. El software utilizado para la captura de imágenes fue digiCamControl 2.0.72.0, el cual permite controlar las cámaras de manera remota en tiempo real desde la PC (foco, distancia, velocidad de obturación, luz, color). En las versiones anteriores de DAL se utilizaba un firmware modificado ([CHDK](#)) que hacía el proceso de captura pero las imágenes se exportaban de la cámara una tarjeta de memoria. Si ocurría algún error en la captura no era posible detectarlo ni modificarlo en el momento. Con el digiCamControl las fotos pasan directamente a ser almacenadas en la computadora. Con el tiempo y el aumento del caudal de materiales que era necesario digitalizar, se hizo perentorio ordenar el proceso y la toma de decisiones técnicas así como a los diferentes actores que intervenían. En la actualidad, la distribución del material a ser escaneado se organiza de la siguiente manera: según el tipo de escaneo a realizar se optará por uno u otro escáner. Para material en buen estado que no esté encuadernado o cuya encuadernación pueda desarmarse conviene utilizar un escáner de alimentación automática (ADF), cuyo desempeño ronda la captura de 60 páginas por minuto (ppm); en nuestro caso el Fujitsu Fi 6170 o el Hp 7500. Cuando el

¹ El índice de reproducción cromática (IRC) es la medida utilizada en relación a una fuente de luz para medir su capacidad de mostrar los colores de un objeto de manera "real". Esto tomando como referencia la iluminación natural.

documento a digitalizar no es plausible de ser desarmado pero su encuadernación permite un ángulo de 180 grados, será necesario un escáner plano, en nuestro caso el HP 7500. Un escáner de grandes dimensiones será necesario para planos y documentos de gran tamaño, en nuestro caso el escáner Context IQ Quattro hasta un ancho máximo de 44 pulgadas o 111,76 centímetros. Un escáner de libros basado en el modelo Archivist 2014 ([«Diseños de escáneres de libros DIY», s.f.](#)) es utilizado para materiales cuya encuadernación requiere un ángulo de apertura menor a 130° y con tamaño de página inferior a 35 cms de alto. Finalmente, el escáner cenital es utilizado con material que presenta un estado delicado de conservación. La disposición de dicho escáner permite sostener libros a 150° aproximadamente, y balancear el peso de ambos lados de modo que la encuadernación no sea forzada. La ausencia de cristal en la cuna del escáner permite una mejor manipulación y un mejor control sobre el movimiento del papel.



Figura 1: Hp 7500 Escaner con alimentacion automatica y cama plana.

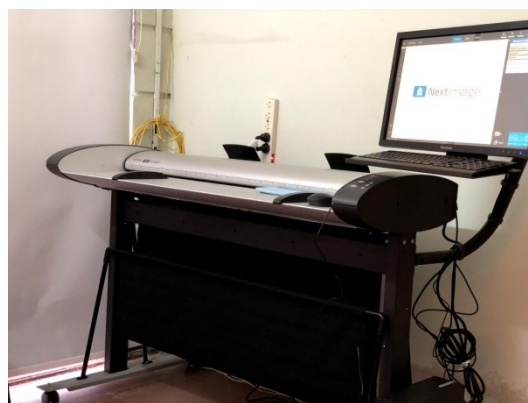


Figura 3: Context IQ Quattro Escáner de gran formato.

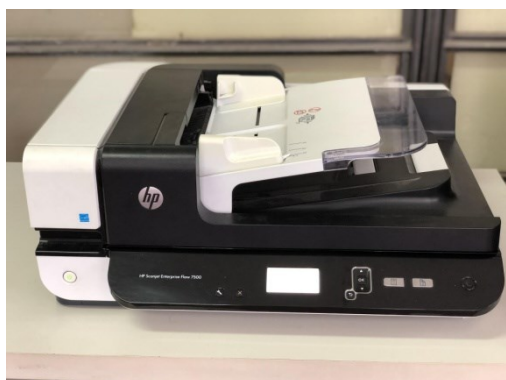


Figura 2: Fujitsu Fi 6170 Escáner de alimentacion automática

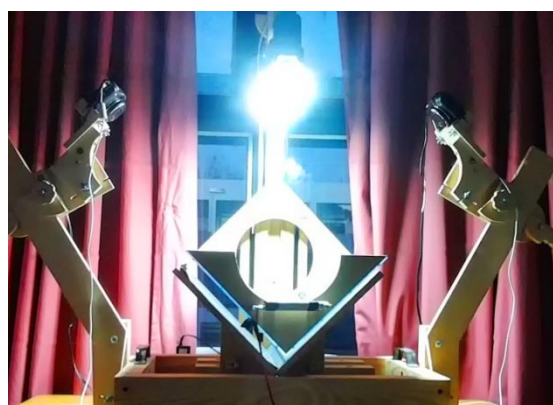


Figura 4: Primer DAL de SEDICI basado en el Model 1 de DIY



Figura 5: Modelo de escáner con câmara cenital

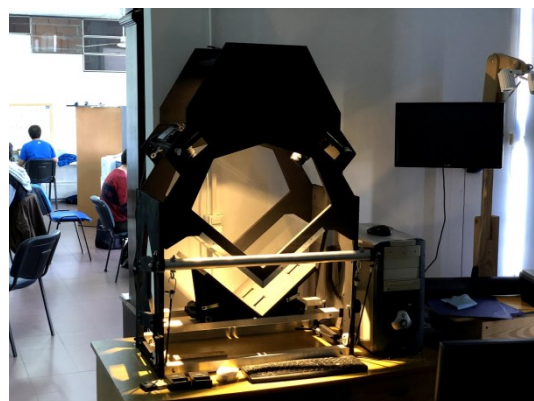


Figura 6: Modelo de escáner Archivist 2014, o Escáner DAL

Flujo de trabajo

Se elaboró un flujo de trabajo que atiende a los diferentes tipos de materiales a ser digitalizados y las particularidades de su digitalización así como un registro de todos los procesos llevados a cabo con el material digital resultante.

Este flujo de trabajo consiste en una serie de etapas (ver gráfico), mayormente consecutivas, que organizan y rigen las acciones a realizar sobre cada material.

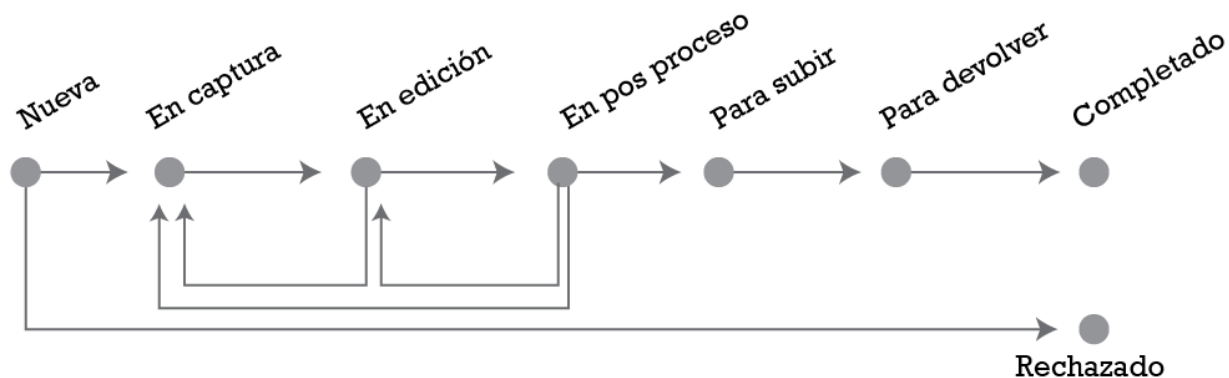


Figura 7: Estados en el proceso de digitalización

Para la gestión de los distintos estados y avances de los materiales ingresados para digitalizar se utiliza Redmine, una herramienta de gestión de tickets. Este sistema permite realizar un seguimiento completo de todas las etapas que atraviesan dichos materiales, los responsables de cada momento y mantiene un registro de las acciones realizadas.

<input type="checkbox"/>	#	Tipo	Estado	Categoría	Prioridad	Asunto	Asignado a	Actualizado	Tipo de recurso	Escáner
■ Nueva 6										
<input type="checkbox"/>	5767	Recurso	Nueva		Normal	Sección II. Labor de los Centros de Estudios. 1940	digitalizadores	24 Mayo 2019 04:42 PM	Documento institucional	DAL
<input type="checkbox"/>	5637	Recurso	Nueva		Normal	Marasso Arturo - Rubén Darío y su creación poética	Pablo Mendez Moura	29 Abril 2019 09:57 AM	Libro	DAL
■ En captura 1										
<input type="checkbox"/>	5771	Recurso	En captura		Normal	Castiñeiras, J.R., Síntesis histórica de las Universidades Argentinas.	Pablo Mendez Moura	31 Mayo 2019 09:09 AM	Documento institucional	DAL
<input type="checkbox"/>	5668	Recurso	Comentarios		Normal	Cremaschi y Tauber - Acceso Ferroviario en Trinchera a la Ciudad de La Plata	Analia Pinto	25 Abril 2019 10:31 AM	Otros	Automático
■ En edición 2										
<input type="checkbox"/>	5751	Recurso	En edición		Normal	UNLP. Digesto nº2. 1935	Pablo Mendez Moura	31 Mayo 2019 01:55 PM	Documento institucional	Plano
<input type="checkbox"/>	5770	Recurso	En edición		Normal	Revista de la Universidad Vol 28	Esteban C Fernandez	31 Mayo 2019 09:44 AM	Boletín/Revista	DAL
■ En postproceso 3										
<input type="checkbox"/>	5760	Recurso	En postproceso		Normal	Sección III. Anuario. 1937	Lorenzo Calamante	31 Mayo 2019 05:23 PM	Documento institucional	DAL
<input type="checkbox"/>	5761	Recurso	En postproceso		Normal	Sección III. Anuario Bibliográfico. 1937.	Lorenzo Calamante	31 Mayo 2019 05:23 PM	Documento institucional	DAL
<input type="checkbox"/>	5762	Recurso	En postproceso		Normal	Sección V. Leyes decretos ordenanzas. 1937	Lorenzo Calamante	31 Mayo 2019 05:23 PM	Documento institucional	DAL
■ Para subir 21										
<input type="checkbox"/>	5741	Recurso	Para subir		Normal	Brelles Mariño, Graciela -Interacción planta-bacteria	sediciano	31 Mayo 2019 06:11 PM	Tesis	DAL
<input type="checkbox"/>	5545	Recurso	Para subir		Alta	Boletín de la asociación de Astronomía Nro14	Carlos J. Nusch	31 Mayo 2019 07:02 AM	Boletín/Revista	Automático
<input type="checkbox"/>	5628	Recurso	Para subir		Normal	Revista de la Universidad Vol 2	Analia Pinto	16 Abril 2019 12:47 PM	Boletín/Revista	DAL
■ Para devolver 22										
<input type="checkbox"/>	5756	Recurso	Para devolver		Normal	Gallo, G., Discursos. 1976	digitalizadores	31 Mayo 2019 11:16 AM	Documento institucional	Plano

Figura 8 - Reporte ejemplo de tareas del proceso de digitalización en Redmine

Etapla 1. Nueva: recepción y ponderación de los materiales

La recepción o revisión del material es un aspecto crucial ya que el repositorio debe evaluar la importancia del material a digitalizar o la premura en digitalizarlo para asegurar la distribución adecuada de la copia física, siempre considerando el flujo total de obras e incluso teniendo en cuenta que algunos materiales por su valor o estado no pueden ser trasladados al repositorio para su digitalización sino que demandan que sean las herramientas del repositorio las que se trasladen para realizar el proceso de digitalización.

Luego se deberá registrar las particularidades de cada material, a saber:

- si el material puede desarmarse o no
- el escáner apropiado de acuerdo al formato,
- persona responsable de realizar digitalización
- la complejidad percibida luego de un análisis preliminar
- los metadatos propios del material (Autor, Título etc)
- los datos del aportante, a quien se le debe devolver el material

Etapla 2. Captura

Para la determinación de esta etapa se utilizaron como insumo diferentes recomendaciones internacionales:

- Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas (IFLA), Consejo Internacional de Archivos (ICA), y Unesco. 2014. «Directrices para proyectos de digitalización de colecciones y fondos de dominio público, en particular para aquellos custodiados en bibliotecas y archivos».
- Junta de Castilla y León. «Recomendaciones para la digitalización de documentos en los archivos».

- Puglia, Steven, Jeffrey Reed, y Erin Rhodes. 2016. «Technical Guidelines for Digitizing Cultural Heritage Materials Creation of Raster Image Files» ed. Federal Agencies Digital Guidelines Initiative (FADGI).

De todos los anteriores la más conveniente fueron las recomendaciones FADGI, en particular para definir resoluciones, aspectos de iluminación y color.

Tipos de materiales ergo tipos de escáneres

El tipo de material tanto como la posibilidad de desarmar el mismo determinarán el escáner que se utilizará para el proceso de digitalización, como se puede observar en el siguiente cuadro. Si el soporte en el que está el material puede ser desarmado o desencuadernado, es posible utilizar los escáneres con alimentación automática que realizan el proceso de digitalización de manera más rápida y eficiente: el escáner Fujitsu FI 7160 o el HP 7500 para hojas hasta tamaño oficio y el Contex IQ Quatro para tamaños mayores.

Si el material a digitalizar no puede ser desarmado, como ocurre generalmente con libros cuyo soporte material reporta un valor en sí mismo, se utilizará un escáner cenital para materiales de tamaño mayor a A4, un escáner de libros fotográfico llamado DAL o un escáner de cama plana HP 7500 plano siempre y cuando el ángulo de apertura de la encuadernación lo permita y no ponga en riesgo el estado general de conservación.



Figura 9: Selección del escáner

Parámetros de captura

De acuerdo al escáner elegido se utilizará un programa de captura diferente. Por ejemplo, en SEDICI usamos:

- digiCamControl para los escáneres de libros y Cenital:
- PaperStream para el escáner automático Fujitsu FI-7160
- NextImage para el escáner de gran formato Context IQ Quattro
- Software de escaneo de documentos inteligente de HP para el escáner automático y plano Hp 7500.

Estos programas deberán ser configurados de acuerdo al material en cuestión y a los parámetros de color, resolución y formato definidos. De forma resumida, se presenta a continuación la tabla de resoluciones utilizada para la captura:

Tipo de material	Profundidad de bit	DPI
Color	24bit	400

Escala de grises	8bit	400
Blanco y negro	2bit	600

Tabla 1: Resoluciones utilizadas por color

Etapa 3. En edición

Para la edición y mejoramiento de imagen se toman los archivos generados en la etapa de captura, y se procesan las imágenes para:

1. Rotar páginas
2. Enderezar las imágenes
3. Ajustar márgenes
4. Eliminar manchas, puntos indeseados
5. Normalizar color
6. Mejorar contraste entre texto y fondo

Para realizar estas acciones se utilizan los productos Scantailor Advanced y Adobe Photoshop CS6.

Etapa 4. de post-proceso

En este punto se consideran distintas tareas que se llevan adelante sobre las imágenes en buen estado, para realizar el reconocimiento óptico de caracteres (OCR) y generar las distintas copias para preservación digital y para la difusión.

OCR

Las imágenes en alta calidad, normalizadas en formato, resolución y color, son incorporadas luego al software de OCR para su procesamiento. Se utiliza un método semi-automático de OCR, que consiste en escanear las imágenes con el software Abbyy FineReader y revisar manualmente las elecciones realizadas por el programa. Desde este software se genera el archivo en formato PDF maestro.

Formatos

Para la preservación digital se guardan los archivos PDF en formato de archivo PDF/A, el cual asegura el acceso y la legibilidad a largo plazo de los archivos por diversos motivos como ser: es un estándar (ISO 19005-1:2005 e ISO 32000-1) y está 100% auto-contenido, lo que implica que con el mero archivo se puede ver todo su contenido sin hacer uso de fuentes o recursos externos. Según las posibilidades del software de transformación se genera PDF/A1 (ISO 19005) o PDF/A-2 (ISO 32000-1). Siempre que sea posible, se generan formatos PDF/A-1a y PDF/A-2a dado que garantizan la accesibilidad, correcta visualización y permiten preservar las características estructurales del documento y el OCR generado. Cuando esto no es posible se genera PDF/A-1b o PDF/A-2b. En casos particulares se utilizan también otros formatos.

El guardado de los archivos

El guardado de los archivos se realiza en un espacio de almacenamiento compartido en la nube. En cada puesto de trabajo se fijan reglas de creación de directorios temporales para que puedan ser ubicados con facilidad por todas las personas que intervienen en el proceso de digitalización. Para esto se crea una partición o unidad de disco separada del sistema operativo.

Se guardan 3 archivos, el primero es un archivo TIFF sin comprimir, segundo un archivo PDF/A sin compresión y por último un archivo PDF/A con compresión. Los primeros dos tipos de archivos de alta calidad de imagen están destinados a la preservación digital y el tercero de ellos, de formato comprimido, se utiliza para la difusión en el repositorio. El último archivo puede implicar la pérdida en la calidad del dato, a fin de lograr archivos de menor tamaño, más amigables para el usuario final y para la indexación por parte de crawlers.

Etapas finales del proceso

Luego del proceso de digitalización se tiene una etapa para reunir al material que debe ser subido al repositorio y otra para organizar el material que resta ser devuelto a su aportante original. Este estado es extremadamente importante ya que todos los materiales antiguos tienen un elevado valor cultural para la institución y deben ser devueltos lo antes posible a su responsable y/o curador. Una vez completados estas etapas se cierra la tarea y se da por completado el trabajo.

Una vez completado el proceso, es posible saber qué materiales se recibieron, quién fue su aportante, cuánto se demoró en cada etapa, cuál fue el proceso de captura, edición y post proceso usado, los archivos resultantes en formato maestro y para difusión, los responsables de cada etapa, fechas, enlace al ítem publicado y enlace al los archivos guardados internamente en la nube, entre muchos otros datos.

Conclusiones

El diseño y aplicación de un flujo de trabajo ordenado y administrado a través de Redmine supuso un mejor aprovechamiento de los recursos y una mayor certeza a la hora de saber qué hardware utilizar, qué metodología de edición tomar y qué software resulta más conveniente de acuerdo al tipo de material a digitalizar. El aumento de número de escáneres y el mejoramiento de los escáneres contruidos por el personal de SEDICI posibilitó un aumento exponencial en el uso de los mismo y la posibilidad de digitalizar un mayor número de documentos en un complejo estado de conservación. En los procesos posteriores a la captura de imagen y edición, también ha permitido una optimización y ordenamiento, sobre todo en la etapa de guardado y preservación de los archivos digitales. La escritura y constante revisión de dichos procesos, redundan en una mejora de los resultados, un mejor aprovechamiento de los recursos y el tiempo empleado en las tareas de digitalización y edición. A su vez, la misma existencia de un Manual de Procedimientos de Digitalización permite al personal contar con un documento al cual recurrir como guía de mejores prácticas y permite, al mismo tiempo, transparentar y dar cuenta de los procesos llevados a cabo con los documentos digitalizados.

Referências

- ABYY *FineReader*. 2013. <https://www.abbyy.com/es-la/?content=global> (3 de junio de 2019).
- Adobe *Photoshop CS6*. 2012. <https://www.adobe.com/la/products/photoshop.html> (8 de agosto de 2018).
- Centro de Servicios en Gestión de Información. 2016. *Manual de procedimientos CIC-Digital*. <http://digital.cic.gba.gob.ar/handle/11746/4562>.
- De Giusti, Marisa Raquel, Silvia Peloché, Carlos Javier Nusch, y Emmanuel Urbina. 2016. «Manual de procedimientos CIC-Digital». <http://digital.cic.gba.gob.ar/handle/11746/4562>.
- «Derecho a Leer (DAL)». <http://derechoaleer.org/blog/acerca-de.html> 3 de junio de 2019).
- digiCamControl*. 2018. <http://digicamcontrol.com/> (3 de junio de 2019).
- «Diseños de escáneres de libros DIY». <http://diybookscanner.org/es/designs.html> (3 de junio de 2019).
- «DIY Book Scanner». <http://diybookscanner.org/es/index.html> (3 de junio de 2019).
- Edmondson, Ray. 2002. «Memoria del Mundo: directrices para la salvaguardia del patrimonio documental».
- Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas (IFLA), Consejo Internacional de Archivos (ICA), y Unesco. 2014. «Directrices para proyectos de digitalización de colecciones y fondos de dominio público, en particular para aquellos custodiados en bibliotecas y archivos». <http://travesia.mcu.es/portaln/jspui/handle/10421/3342> (3 de junio de 2019).
- Foster, Stephen, Roslyn Russell, Jan Lyall, y Duncan Marshall. 1995. «Directrices para la salvaguardia del patrimonio documental». https://unesdoc.unesco.org/ark:/48223/pf0000105132_spa (23 de junio de 2019).
- Hewlett-Packard Co. «Escáner plano HP Scanjet Enterprise Flow 7500 - Descargas de controladores | Soporte al cliente de HP®». <https://support.hp.com/ar-es/drivers/selfservice/hp-scanjet-enterprise-flow-7500-flatbed-scanner/5448147> (3 de junio de 2019).
- International Organization for Standardization (ISO). 2015. «UNE-ISO 14721:2015 Sistemas de transferencia de datos e información espaciales. Sistema abierto de información de archivo (OAIS). Modelo de referencia.» <https://www.aenor.com/normas-y-libros/buscador-de-normas/iso/?c=062542> (22 de abril de 2019).

- . 2016. «ISO 15489-1:2016 Information and documentation -- Records management -- Part 1: Concepts and principles». <https://www.aenor.com/normas-y-libros/buscador-de-normas/iso/?c=062542> (22 de abril de 2019).
- . 2017. «UNE-ISO 16363:2017 Sistemas de transferencia de información y datos espaciales. Auditoría y certificación de repositorios digitales de confianza». <https://www.aenor.com/normas-y-libros/buscador-de-normas/iso/?c=062542> (22 de abril de 2019).
- Junta de Castilla y León. «Recomendaciones para la digitalización de documentos en los archivos». <https://archivoscastillayleon.jcyl.es/web/jcyl/ArchivosCastillaYLeon/es/Plantilla100Detalle/1253861401355/1253861401355/1284166105811/Comunicacion> (3 de junio de 2019).
- Library of Congress. «Technical Information». *Technical Information, Library of Congress*. <https://memory.loc.gov/ammem/about/techIn.html> (3 de junio de 2019).
- Muñoz, Wilson Castaño, y María Camila Restrepo. 2016. «Los repositorios como herramienta para la recuperación del patrimonio bibliográfico: el caso de seis bibliotecas públicas municipales del departamento de Antioquia». *Revista Interamericana de Bibliotecología* 39(1): 57-68.
- «Nextimage Software». *Contex*. <https://contex.com/support/software-nextimage/> (3 de junio de 2019).
- «Nuestro Do It Yourself Book Scanner». <http://derechoaleer.org/blog/2011/08/nuestro-do-it-yourself-book-scanner.html> (3 de junio de 2019).
- Nusch, Carlos et al. 2018a. «Digitalización y Preservación Digital del patrimonio museológico de la Universidad Nacional de La Plata: Copiador – Facultad de Ciencias Físicas, Matemáticas y Astronómicas (1918-1925)». <http://sedici.unlp.edu.ar/handle/10915/69958> (3 de junio de 2019).
- . 2018b. «Particularidades y desafíos en la digitalización del libro Copiador – Facultad de Ciencias Físicas, Matemáticas y Astronómicas (1918-1925): un caso de uso de un repositorio institucional para la preservación de Memoria Institucional». <http://sedici.unlp.edu.ar/handle/10915/70355> (3 de junio de 2019).
- «PaperStream Capture - Fujitsu Spain». <https://www.fujitsu.com/es/products/computing/peripheral/scanners/fi/software/ps-capture/ps-capture.html> (3 de junio de 2019).
- Pérez Pineda, Iván. 2017. «Plan de preservación digital del patrimonio documental digitalizado de la Biblioteca Lafragua». <http://diposit.ub.edu/dspace/handle/2445/117724> (17 de abril de 2019).



Universidad Nacional de La Plata, Instituto de Física La. 1918. *Libro Copiador de la Facultad de Ciencias Físicas, Matemáticas y Astronómicas (1918-1925)*. <http://hdl.handle.net/10915/67640> (3 de junio de 2019).

Puglia, Steven, Jeffrey Reed, y Erin Rhodes. 2016. «Technical Guidelines for Digitizing Cultural Heritage Materials Creation of Raster Image Files» ed. Federal Agencies Digital Guidelines Initiative (FADGI).

ScanTailor. 2012. <http://scantailor.org/> (8 de agosto de 2018).

HUB-UR: Services and Expert Finder - una solución de visibilidad web de la Universidad del Rosario – Colombia

Humberto Blanco-Castillo¹, Malgorzata Lisowska²

¹ Universidad del Rosario. Correo: humberto.blanco@urosario.edu.co

² Universidad del Rosario. Correo: margarita.lisowska@urosario.edu.co

Resumen

La visibilidad web, en la actualidad, es uno de los temas prioritarios para muchas instituciones de educación superior y es por eso que desde las diferentes áreas de una universidad (biblioteca, investigación, informática, etc.) constantemente se diseñan estrategias tendientes a su fortalecimiento; e incluso se implementan herramientas robustas. En general, es difícil encontrar una plataforma capaz de integrar en un solo lugar toda la información institucional, aunque frecuentemente se crean diferentes subdominios web, cada uno enfocado a un propósito específico (páginas web generales, blogs, repositorios institucionales, CRIS, etc.), con una información aislada y no vinculada. Por lo consiguientes es muy difícil, por no decir casi imposible construir servicios utilizando la información institucional de forma globalizada e interoperable, como por ejemplo buscar un experto o un tutor de tesis. Con el fin de contar con una herramienta de visibilidad web realmente integral y que permita crear servicios basados en las capacidades individuales e institucionales de la Universidad del Rosario, el Centro de Recursos para el Aprendizaje y la Investigación – CRAI ha implementado una solución que integra diferentes fuentes de información institucional. Esta ponencia pretende describir el proceso de definición de la estrategia, implementación y lanzamiento de la herramienta llamada **HUB-UR: Services and Expert Finder**, basada en el software de código abierto VIVO y cuyos beneficios de la web semántica, permiten visibilizar de forma estructurada, vinculada, ágil y en un solo lugar, las fortalezas de la Universidad del Rosario de Colombia.

Palabras clave: Software VIVO, Visibilidad web, Interoperabilidad, web semántica.

Introducción

El Centro de Recursos para el Aprendizaje y la Investigación – CRAI de la Universidad del Rosario, ha sido líder, en Colombia y en la región, en la promoción de acceso abierto y en el aprovechamiento de sus beneficios para diseñar estrategias que permitieran fortalecer la visibilidad institucional en la web. Es así que ha creado, desde el año 2008, el Repositorio Institucional EdocUR, que hoy cuenta con más de 3 millones de visitas al año, y ha sido galardonado en diferentes

ocasiones. En al año 2010 lideró el proceso de implementación del portal de revistas institucionales, basado en software libre Open Journal System – OJS, que permitió cambiar la producción de revistas institucionales, del papel a la web. Como primera Universidad en Colombia, en 2016 ha implementado la Política Institucional de Acceso Abierto y en 2017 ha firmado la Declaración de Berlín. Colciencias en 2017 le ha otorgado el Primer Premio a la Ciencia Abierta, que facilitó la expedición de la Política Institucional de Datos de Investigación y el Repositorio de Datos de Investigación. La Universidad también fue el escenario de la primera edición de la Conferencia Internacional sobre Repositorios y Bibliotecas Digitales- BIREDIAL 2011.

Justificación

Con la transformación de biblioteca al CRAI, en 2017, se han diseñado una serie de servicios personalizados para los investigadores encaminados al fortalecimiento de su presencia web. De esta manera se crearon y/o actualizaron los perfiles web del 98% de los investigadores core de la Universidad, principalmente Google Scholar, Scopus ID, Iralis, ORCID, REDALYC.

Para continuar construyendo una estrategia institucional de visibilidad web, era necesario pensar en la implementación de una herramienta integral que permitiera concentrar en un solo lugar la información de calidad y actualizada sobre las capacidades y fortalezas académicas e investigativas de la Universidad del Rosario, garantizando integración de los sistemas de información, agilidad en la recuperación y usabilidad. Cada vez tomaba más fuerza la necesidad de crear servicios, basados en las fortalezas y capacidades de la Universidad del Rosario, para los siguientes grupos de usuarios, internos y externos:



Gráfico 1: Grupos de usuarios

Definición de la herramienta

En la búsqueda de encontrar una herramienta adecuada hemos revisado algunas soluciones existentes, tanto de pago, como abiertas. Finalmente nos hemos decidido por el software libre VIVO, debido a su flexibilidad, interoperabilidad, nivel de adopción y la comunidad que lo soporta. VIVO es una plataforma de código abierto basada en web semántica, desarrollada por la Universidad de Cornell y soportada por la comunidad Duraspace.

La arquitectura de un sistema VIVO interrelaciona diferentes capas como:

- ✓ **Capa de presentación** cuya función es filtrar que tipo de contenido debe ser expuesto, junto con plantillas freemaker para la visualización de todas las entidades.
- ✓ **Capa de lógica de negocio**, que contiene las reglas de negocio para el acceso, las ontologías, el motor de consultas a través de SPARQL, entre otros.
- ✓ **Capa de persistencia** donde se almacenan las tripletas en la base de datos y el motor de SOLR encargado de las búsquedas, filtros y facetas.



Gráfico 2: Arquitectura del sistema VIVO

VIVO utiliza un motor de web semántica (vitro) y una colección de ontologías para representar los datos almacenados en su base de datos.

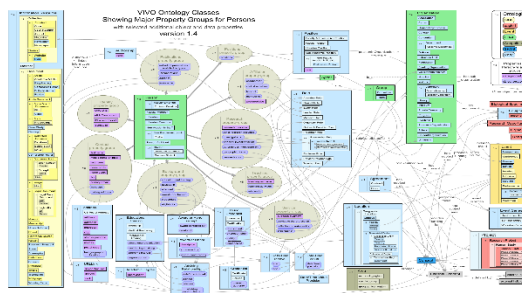


Gráfico 3: Ontología VIVO

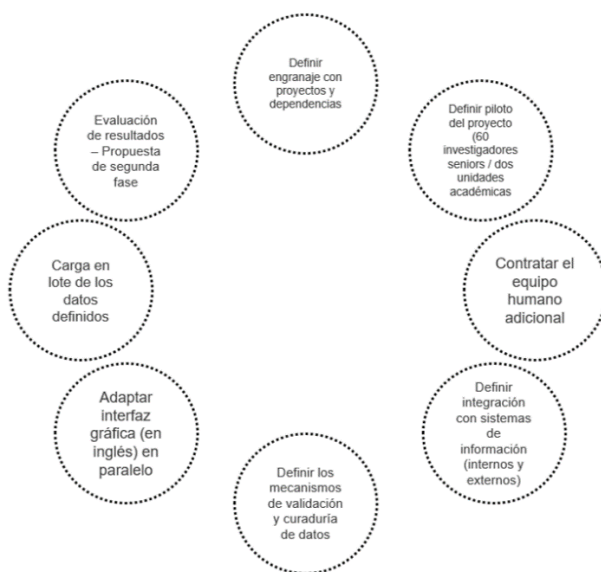
VIVO utiliza Linked Open Data, para relacionar y vincular la información de perfiles individuales, institucionales, proyectos, productos de investigación, eventos, etc; provenientes de diferentes fuentes institucionales.

Las principales fortalezas de VIVO están relacionadas con los siguientes asuntos:

- Ofrecimiento de resultados robustos con búsquedas simples.
- Exposición de información interconectada e interrelacionada.
- Visualización de los perfiles de investigadores, que incluye las gráficas de coautoría y coinvestigación.
- Es robusto y escalable, pudiendo integrarse como página principal de otros portales.

Desarrollo

Una vez estructurado el proyecto inicial por el CRAI, se involucraron otras áreas de la Universidad que gestionan la información académica, de investigación y extensión, quienes robustecieron el alcance. La propuesta final del proyecto, se presentó de manera conjunta para la aprobación de la Vicerrectoría Académica, y se obtuvo la financiación interna para implementarlo.



Gráfica 4: Etapas del proyecto

La implementación del **HUB-UR: Services and Expert Finder** se desarrolló por fases, que se describen a continuación:

Fase 1 – Diagnóstico de necesidades

Dado que **HUB-UR: Services and Expert Finder** es un portal público, se revisó cuidadosamente la información a visualizar, además de la integración con otros portales ya existentes. Por ejemplo, en una rápida revisión de las fuentes institucionales, que contenían alguna información relacionada con los investigadores, encontramos que existían las siguientes posibilidades:

- ✓ **Sistema de gestión de investigación** con la producción científica y los proyectos del investigador.
- ✓ **Sistema de gestión docente**, que incorpora información sobre intereses, capacidades y el último nivel de estudio de un investigador.
- ✓ **Registros de las actividades de extensión**, en ausencia de un sistema unificado.
- ✓ **Repositorio institucional**, que contiene las asesorías de trabajo de grado y la información de eventos de visibilidad institucional.
- ✓ **Directorio de profesores (investigadores)** de la página web institucional.

Algunos de estos recursos se gestionan de forma manual, por lo que frecuentemente la información no se encuentra sincronizada, ni actualizada.

De acuerdo a los resultados de la discusión del grupo interdisciplinario institucional acerca de enfoques, limitaciones y representaciones del Hub, se estableció que se presentará la información de acuerdo a los siguientes agrupamientos: perfiles de investigadores, unidades académicas, producción científica, proyectos de investigación y eventos de visibilidad institucional.

Para conocer el rol de Hub y sus diferentes niveles de interacción con otros portales de visibilidad institucional, como PURE, página web, etc., se diseñó de manera preliminar el mapa de sistemas de información UR. Esto sirvió como insumo para la generación del proyecto de automatización para extraer la información de los sistemas fuente y su carga en el HUB-UR.

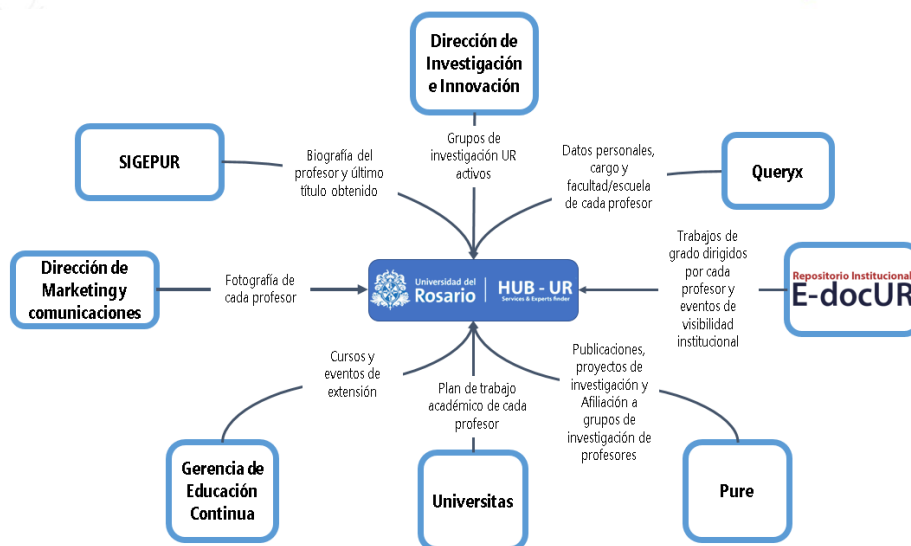


Gráfico 5: Integración con sistemas de información fuente

Finalmente se definió el siguiente alcance:

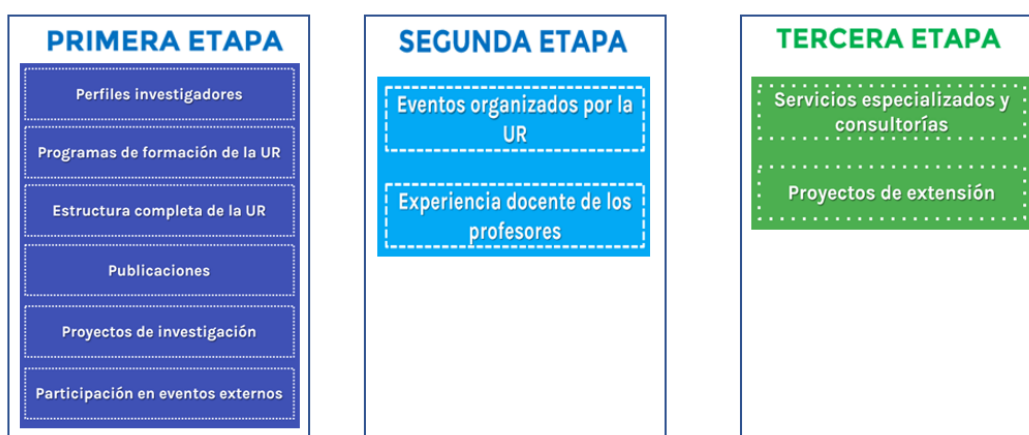


Gráfico 6: Fases de implementación

Fase 2 – Implementación

Definidas las necesidades de información de Hub y la interoperabilidad con otras herramientas institucionales, se procedió con su implementación.

Se realizó el despliegue de un ambiente de prueba con la versión 1.9 de VIVO y se desarrolló una prueba piloto, con el fin de:

- Conocer las características de los perfiles y objetos en VIVO.
- Definir el orden de ingreso de información.
- Entender las diferencias de información de los objetos en VIVO.

- Definir los requerimientos de información para cada etapa del proyecto.

Posteriormente, se establecieron acuerdos y planes de acción con el equipo de trabajo responsable de la gestión de los sistemas de información fuente (Dirección académica, Dirección de tecnología informática y de comunicaciones, Dirección de Investigación e Innovación, Dirección de gestión humana, Dirección de tecnología, Dirección de Extensión), con el fin de:

- Normalizar y validar la información (Limpieza de datos, desambiguación y modelado semántico).
- Establecer requerimientos de interoperabilidad entre sistemas.
- Definir la periodicidad de la actualización de la información recolectada.
- Determinar los mecanismos para la generación de datos institucionales, no existentes en la actualidad en los sistemas fuentes, pero relevantes para la visualización de la información en VIVO, como coautorías externas.

Una vez finalizada la fase de extracción, transformación y carga de información en el servidor de pruebas, se realiza la verificación de la coherencia y visualización de datos.

Al mismo tiempo, otro equipo se encargó del aspecto gráfico de la plataforma, con el fin de añadir soporte responsive, la cual no se encontraba disponible totalmente en la versión 1.9 y buscando también mejorar la experiencia del usuario final.

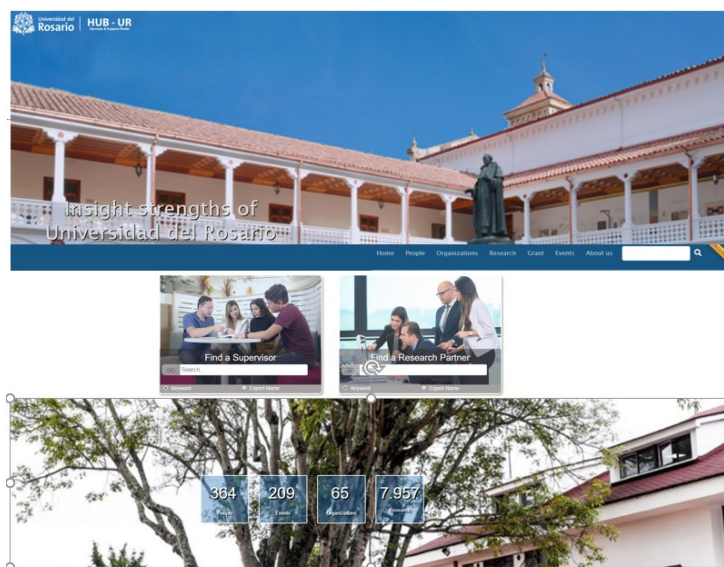


Gráfico 7: Home page HUB-UR

Una vez aprobados, el diseño gráfico y la calidad de la información presentada, se realiza la migración al servidor de producción y se asigna el dominio research-hub.urosario.edu.co.

Servicios

Uno de los factores diferenciadores que destacan el HUB-UR es la presentación de las fortalezas institucionales a través de servicios, con el fin de facilitar la consulta y recuperación de información de acuerdo a los intereses del usuario. Estos servicios incluyen filtros, para que una vez se haya realizado la búsqueda, se puedan refinar los resultados de acuerdo a criterios específicos (por ejemplo, área de experiencia o nivel de formación).

En la actualidad, la plataforma cuenta con 2 servicios implementados: *Find a supervisor* (para encontrar un tutor de tesis) y *Find a Research partner* (para encontrar un socio de investigación). Sin embargo, para este año se tiene planeado el diseño e implementación de 4 servicios más y se espera que la generación de servicios sea un proceso constante teniendo en cuenta la información disponible y las necesidades de los usuarios.

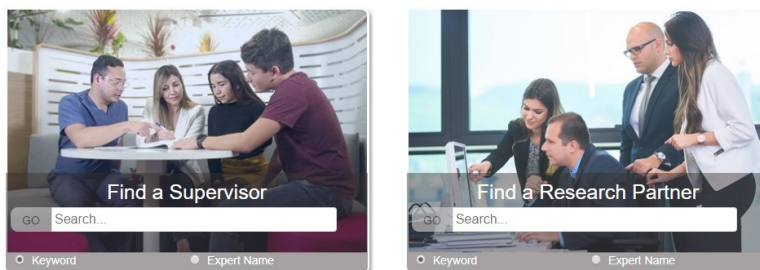


Gráfico 8: Servicios HUB-UR

Fase 3 – Lanzamiento

Dentro de la fase de lanzamiento, se determinó que un factor clave para éxito de la plataforma es su divulgación; por esta razón se diseñó una estrategia dirigida a estudiantes como potenciales usuarios y a investigadores como protagonistas y beneficiarios de la visibilidad académica en la web.

La implementación de la estrategia de comunicación se realizó en tres etapas:

1. **Campaña de expectativa:** en esta etapa, se crearon 2 videos de expectativa, uno dirigido a estudiantes y otro a investigadores, los cuales se publicaron una semana antes del lanzamiento a través de redes sociales para el caso de los estudiantes y de correo electrónico para los investigadores .
2. **Día del Lanzamiento:** Se escogió el día del profesor como la mejor fecha para realizar el lanzamiento, considerando que este día es posible tener la mayor concentración de profesores en un mismo auditorio. Para el lanzamiento, se realizó un video en el cual se presenta su propuesta de valor y los elementos que hacen del HUB-UR una plataforma de gran importancia para la visibilidad académica en la web; adicionalmente, se entregó un objeto promocional personalizado con la visualización del perfil del investigador en

el HUB-UR y un código QR que enlaza directamente con el perfil del investigador.

3. **Remarketing:** posteriormente al lanzamiento se ha planeado enviar mensajes dirigidos para reforzar la importancia del HUB-UR como aliado de la visibilidad académica web.

Desafíos

Calidad de la información: La integración de la información de los sistemas fuentes en el HUB-UR ha permitido evidenciar inconsistencias en los datos, lo que ha planteado la necesidad de realizar actividades conjuntas entre los responsables de la gestión de dichas plataformas para estandarizar la información reportada en los diferentes sistemas.

Interoperabilidad de los sistemas de información: A partir del proyecto de HUB-UR, ha surgido la necesidad de generar interfaces entre los distintos sistemas de información de la universidad, para optimizar la disponibilidad de la información y evitar los errores generados en el procesamiento manual.

Validación y normalización de la información: Se requiere la validación de toda la información que será divulgada en el HUB-UR, debido a la importancia y el impacto que se espera generar por parte de la plataforma.

Adición de nueva información: identificación de nuevas fuentes de información para que permitan visualizar otras fortalezas de la universidad en docencia, investigación y extensión, con el fin de generar nuevos servicios.

Posicionamiento del HUB-UR: Es necesario realizar todas las actividades de posicionamiento WEB requeridas para que la plataforma se constituya como la primera fuente de búsqueda de investigadores o fortalezas de la universidad.

Procedimientos alternos de actualización de información: Mientras se automatizan los procesos de sincronización de la información entre los diferentes sistemas fuente, los procedimientos de actualización de información deben garantizar la coherencia de la información cargada en el HUB-UR

Conclusiones

La expansión de la Web y su importancia en el contexto académico permite que hoy cada institución de educación superior, esté en la capacidad de poner a disposición del mundo, sus fortalezas en términos de investigación, docencia y extensión. Sin embargo, esta posibilidad ha hecho que esta información se presente desde distintos portales de la misma institución causando pérdida de efectividad en la recuperación de información o ambigüedad en los resultados, afectando la reputación de las instituciones.

El HUB de fortalezas se constituye como un proyecto relevante para la institución, por su carácter integrador, lo que exige la participación de los responsables de los sistemas fuente trabajando en estrecha colaboración con el fin de garantizar la calidad de la información del HUB-UR. La utilidad se ve reflejada



en la integración de la información de diferentes fuentes para visualizar en un solo lugar publicaciones de los investigadores, sus proyectos y otras actividades de docencia y extensión en una interfaz liviana e interoperable.

Este tipo de proyectos integradores permite depurar la información de los diferentes sistemas fuente, encontrando incoherencias de información, lo que contribuye al mejoramiento de los sistemas de información institucionales.

El HUB es una herramienta que fortalece la visibilidad académica en la web, mostrando las áreas de experticia, credenciales académicas, redes de trabajo, publicaciones, proyectos, eventos y capacidades institucionales.

Facilita la colaboración interdisciplinaria entre miembros de diferentes facultades y escuelas y la conexión entre diferentes actores internos y externos como, por ejemplo, un estudiante y su potencial tutor de tesis, un investigador y sus pares, un ente externo y los grupos de investigación

En la actualidad, se han recibido más de 24 mil visitas de distintos países¹ lo que permite ver un panorama alentador para el futuro de la plataforma.

¹ 74,34% son de Colombia, un 17,5% de Estados Unidos y el 8% restante se encontró distribuido entre usuarios de Perú, Chile, México, España, Ecuador, Canadá, Brasil y Alemania

LUME - Módulo de estatísticas 2.0 para o Dspace 5.8

Cleusa Pavan¹; Manuela Klanovicz Ferreira²; André Rolim Behr³; Caterina Marta Groposo Pavão⁴; Janise Silva Borges da Costa⁵

¹ Universidade Federal do Rio Grande do Sul. Email: cleusa.pavan@cpd.ufrgs.br

² Universidade Federal do Rio Grande do Sul. Email: manuelakf@cpd.ufrgs.br

³ Universidade Federal do Rio Grande do Sul. Email: andre.behr@cpd.ufrgs.br

⁴ Universidade Federal do Rio Grande do Sul. Email: caterina@cpd.ufrgs.br

⁵ Universidade Federal do Rio Grande do Sul. Email: janise@cpd.ufrgs.br

Resumo

O monitoramento e a análise dos padrões de utilização dos repositórios institucionais permitem aos gestores avaliar e aprimorar a ferramenta. Os mantenedores de repositórios têm, dentre outras, a função primordial de aumentar a visibilidade de sua produção científica e, conseqüentemente, ampliar a leitura, a citação e a socialização do conhecimento. O trabalho descreve o Lume - Módulo de estatísticas 2.0, desenvolvido com base na versão disponível no DSpace 5.8, que permite a visualização gráfica dos dados de acessos e *downloads*, não disponível na versão original do DSpace. As estatísticas implementadas foram: a) número de *downloads* de todo o repositório, discriminados por comunidade, subcomunidade e coleção; b) os dez itens com maior número de *downloads* em cada subcomunidade de segundo nível; c) acesso e *downloads* agrupados por comunidade, subcomunidade, coleção ou item; d) por autor ou assunto, ambas permitindo o filtro de determinado período com visualização de gráficos por ano ou por mês, por top 10 países ou por todos os países de origem. Aborda, também, como adicionar o Módulo ao DSpace 5.8, a partir de um repositório público de código fonte.

Palavras-chave: Repositórios institucionais. DSpace 5.8. Estatísticas. Visibilidade.

Abstract

Monitoring and analyzing the usage patterns of institutional repositories allows managers to evaluate and improve the tool. Repository maintainers have, among others, the prime function of increasing the visibility of their scientific production. Consequently, increasing the reading, quotation, and socialization of knowledge. The paper describes Lume - Statistics Module 2.0, developed based on the version available in DSpace 5.8, which allows graphical visualization of access and download data, not available in the original version of DSpace. The statistics implemented were: a) number of downloads of the entire repository, broken down

by community, subcommunity, and collection; b) the ten items with the highest number of downloads in each second level subcommunity; c) access and downloads grouped by community, subcommunity, collection, or item; d) by author or subject, both allowing the filter of a defined period with graph visualization by year or by month, by top 10 countries or by all countries of origin. It also covers how to add the Module to DSpace 5.8, from a public repository of source code.

Keywords: Institutional repositories. DSpace 5.8. Statistics. Visibility.

Introdução

Os repositórios institucionais (RI) em acesso aberto, ao armazenarem e disponibilizarem a produção científica das instituições de ensino e pesquisa colaboram para aumentar a sua visibilidade e o seu uso. A visibilidade é um dos indicadores utilizados pelos sistemas de avaliação para determinar a posição de prestígio de um pesquisador, uma instituição ou um país. De acordo com Leite (2009), a contribuição principal dos RI está na reformulação e melhoria da comunicação científica ao adotar processos de gestão da informação científica.

Além desses aspectos, os RI fornecem dados relativos à sua utilização, podendo evidenciar o comportamento dos usuários em relação à ferramenta e ao conteúdo disponibilizado. Ao fazer *download* do texto, e não apenas visualizar os resultados de busca, o usuário manifesta interesse pelo documento recuperado. Por outro lado, estas informações também podem impulsionar os gestores dos RI a aprimorar e desenvolver novas funcionalidades.

O DSpace é a plataforma mais empregada pelos RI, com 43% do total de 4.140, conforme contabiliza o OpenDOAR (DIRECTORY OF OPEN ACCESS REPOSITORIES, 2019). A segunda posição é ocupada por EPrints com 13% e as demais plataformas são usadas por 3% ou menos dos repositórios.

O Lume, repositório digital da Universidade Federal do Rio Grande do Sul (UFRGS), foi desenvolvido utilizando o DSpace e implantado em 2008. Em maio de 2019, disponibilizava um pouco mais de 211.000 itens, abrangendo todas as áreas do conhecimento.

Desde 2011, dispunha de um módulo de estatísticas customizado com novas formas de visualização de acessos e *downloads*, onde as informações eram armazenadas e consultadas em uma tabela do banco de dados relacional, no caso o PostgreSQL. Conforme aumentava a quantidade de dados armazenados na tabela, a visualização passou a apresentar baixa *performance*. Mesmo com modificações que previam o pré-processamento de consultas, resumindo os dados antecipadamente, e a implantação de *cache*, não foi possível alcançar uma *performance* desejável.

A partir disso e, preparando a migração do Lume para o DSpace 5.8, em 2018, analisou-se a forma como os dados eram armazenados e consultados nesta versão. Constatou-se que o registro era feito no SOLR, uma plataforma de pesquisa e base de dados documental que permite consultas complexas aos seus

dados. Esta versão do DSpace disponibiliza, originalmente, três tipos de estatísticas: de uso, de busca e de fluxo de submissão, sem representação gráfica. As estatísticas podem ser visualizadas para todo repositório, por comunidade, subcomunidade, coleção ou item, e permitem verificar a quantidade de acessos dos últimos sete meses, por país e por cidade.

O Lume - Módulo de estatísticas 2.0, foi desenvolvido com o objetivo de gerar novas opções de agrupamento e visualização dos dados, sob a forma de tabelas e gráficos, mediante consulta no registro *default* de estatísticas do DSpace 5.8, facilitando a leitura. Além disso, visa sanar o problema de desempenho do módulo implantado em 2011.

Implementação técnica do Módulo

Após a análise e a realização de testes da versão original do DSpace 5.8, modificou-se o antigo módulo de estatísticas do Lume para que consultasse a interface SOLR, disponível no DSpace 5.8, ao invés do banco de dados PostgreSQL. Os primeiros resultados não alcançaram a *performance* esperada, mesmo tendo a consulta mais rápida que no anterior. Realizados novos estudos e testes, foram trabalhados dois pontos:

- a) a consulta feita por período aos dados no SOLR estava muito demorada, então decidiu-se criar índices resumidos por ano e por mês;
- b) a consulta por facetas, disponível no SOLR, foi otimizada de 10 segundos para 1.5 quando utilizado o parâmetro de consulta do SOLR `facet.query`, ao invés do parâmetro `facet.field`, conforme documentado em pergunta no fórum StackOverflow¹.

Ao atingir a *performance* desejada, foi feito o *release* com as seguintes modificações necessárias para implementar o Lume - Módulo de estatísticas 2.0:

- a) alteração do `<dspace-source>/dspace/solr/statistics/schema.xml` com adição de dois campos resumidos de data por ano e por mês, e reindexação do *core* de estatísticas do SOLR (Figura 1);
- b) inclusão e alteração de arquivos fonte em:
 - JAVA: para a consulta das estatísticas do SOLR e a geração de DRI foi incluída uma classe JAVA para cada tipo de visualização das estatísticas;
 - JAVASCRIPT: para a geração dos gráficos foi utilizada a biblioteca `chart.min.js`² e também foi criado o arquivo `chartEstatisticas.js`;
 - CSS: foram incluídas regras para o posicionamento das estruturas nas páginas de visualização.
- c) modificação no arquivo `<theme>/xsl/core/navigation.xsl` de menu lateral para incluir o *link* para as estatísticas nas respectivas páginas;

¹ <https://stackoverflow.com/questions/51247635/debug-a-solr-query-with-facet-field-shows-facet-result-over-more-documents-than>

² <http://chartjs.org/>

```
<!--Adicionando para conter a data apenas com ano e mes-->
<fieldType name="keywordTimeFilterAnomes" class="solr.TextField" sortMissingLast="true" omitNorms="true">
  <analyzer>
    <charFilter class="solr.PatternReplaceCharFilterFactory"
      pattern="(^\d{4}-\d{2})(.*)" replacement="$1"/>
    <!--Treats the entire field as a single token, regardless of its content-->
    <tokenizer class="solr.KeywordTokenizerFactory"/>

    <!--<filter class="solr.LowerCaseFilterFactory" />-->
    <filter class="solr.TrimFilterFactory" />
  </analyzer>
</fieldType>

<!--Adicionando para conter a data apenas com ano e mes-->
<fieldType name="keywordTimeFilterAno" class="solr.TextField" sortMissingLast="true" omitNorms="true">
  <analyzer>
    <charFilter class="solr.PatternReplaceCharFilterFactory"
      pattern="(^\d{4})(.*)" replacement="$1"/>
    <!--Treats the entire field as a single token, regardless of its content-->
    <tokenizer class="solr.KeywordTokenizerFactory"/>

    <!--<filter class="solr.LowerCaseFilterFactory" />-->
    <filter class="solr.TrimFilterFactory" />
  </analyzer>
</fieldType>

[...]

<!--adicionado para diminuir o tempo-->
<field name="time_anomes" type="keywordTimeFilterAnomes" />
<field name="time_ano" type="keywordTimeFilterAno" />
<copyField source="time" dest="time_anomes" />
<copyField source="time" dest="time_ano" />
```

Figura 1: Alteração do arquivo <dspace-source>/dspace/solr/statistics/schema.xml</>

d) modificação do sitemap.xmap para suportar os *links* para as estatísticas (Figura 2).

O Módulo aproveita os registros *default* de estatísticas coletadas pelo DSpace 5.8, através de consulta à interface SOLR, o que minimiza o número de alterações necessárias para a sua implantação e permite que as estatísticas originais permaneçam ativas. Além disso, permite que, no futuro, sejam pensadas novas formas de visualização para as estatísticas de busca e fluxo de submissão, que também são registradas por *default* no SOLR.

Ao registrar as estatísticas *default*, o DSpace 5.8 identifica quais são resultantes das buscas automáticas feitas no Repositório pelos robôs dos motores de busca, a fim de indexar o seu conteúdo. Essas estatísticas ficam marcadas com "isBot:true" e não são contabilizadas na exibição para o usuário final.

O *download* do Módulo e a identificação das modificações necessárias estão disponíveis no GitHub.

```
<map:sitemap xmlns:map="http://apache.org/cocoon/sitemap/1.0">
  <map:components>
    <map:transformers>
      <map:transformer name="BrowseStats" src="org.dspace.app.xmlui.aspect.statistics.BrowseStats"/>
      <map:transformer name="Downloads" src="org.dspace.app.xmlui.aspect.statistics.Downloads"/>
      <map:transformer name="Stats" src="org.dspace.app.xmlui.aspect.statistics.Stats"/>
      <map:transformer name="TopTen" src="org.dspace.app.xmlui.aspect.statistics.TopTen"/>
    </map:transformers>
  </map:components>
  <map:pipelines>
    <map:pipeline>
      <map:generate/>

      <map:match pattern="stats/downloads">
        <map:transform type="Downloads"/>
      </map:match>

      <map:match pattern="stats/topten">
        <map:transform type="TopTen"/>
      </map:match>

      <map:match pattern="browsestats">
        <map:transform type="BrowseStats"/>
      </map:match>

      <map:match pattern="handle/*/*/browsestats">
        <map:transform type="BrowseStats"/>
      </map:match>

      <map:match pattern="handle/*/*/stats">
        <map:transform type="Stats"/>
      </map:match>

      [...]
      <!-- Not a URL we care about, so just pass it on. -->
      <map:serialize type="xml"/>
    </map:pipeline>
  </map:pipelines>
</map:sitemap>
```

Figura 2: Alteração do arquivo sitemap.xmap

As estatísticas de todos os níveis do Lume estão disponíveis para o público sem a necessidade de *login*. Podem ser consultadas por comunidade, subcomunidade, coleção ou item e são identificadas por meio de ícone próprio:



Exibição das estatísticas

As estatísticas são apresentadas a partir da aba Sobre, opção Estatísticas gerais ou vinculadas a cada nível do Repositório, autor e assunto.

Para fins de ilustração dos exemplos apresentados nas alíneas a e b, adotou-se a comunidade Produção científica.

O módulo disponibiliza as seguintes estatísticas:

a) número de *downloads*

O total geral de *downloads*, sob a forma de lista, é exibido numa única página, bem como o número de *downloads* por comunidade, subcomunidade e coleção (Figura 3).

É possível ter um panorama de *downloads*, de acordo com a organização das comunidades adotadas pelo Lume, que abrangem órgãos específicos da UFRGS, programas de pós-graduação, áreas do conhecimento e tipos de materiais. Este número, combinado com outros dados que podem ser extraídos pelo administrador da plataforma DSpace, a partir de tabelas de logs, viabiliza análises e estudos métricos, contribuindo para a gestão da produção institucional.

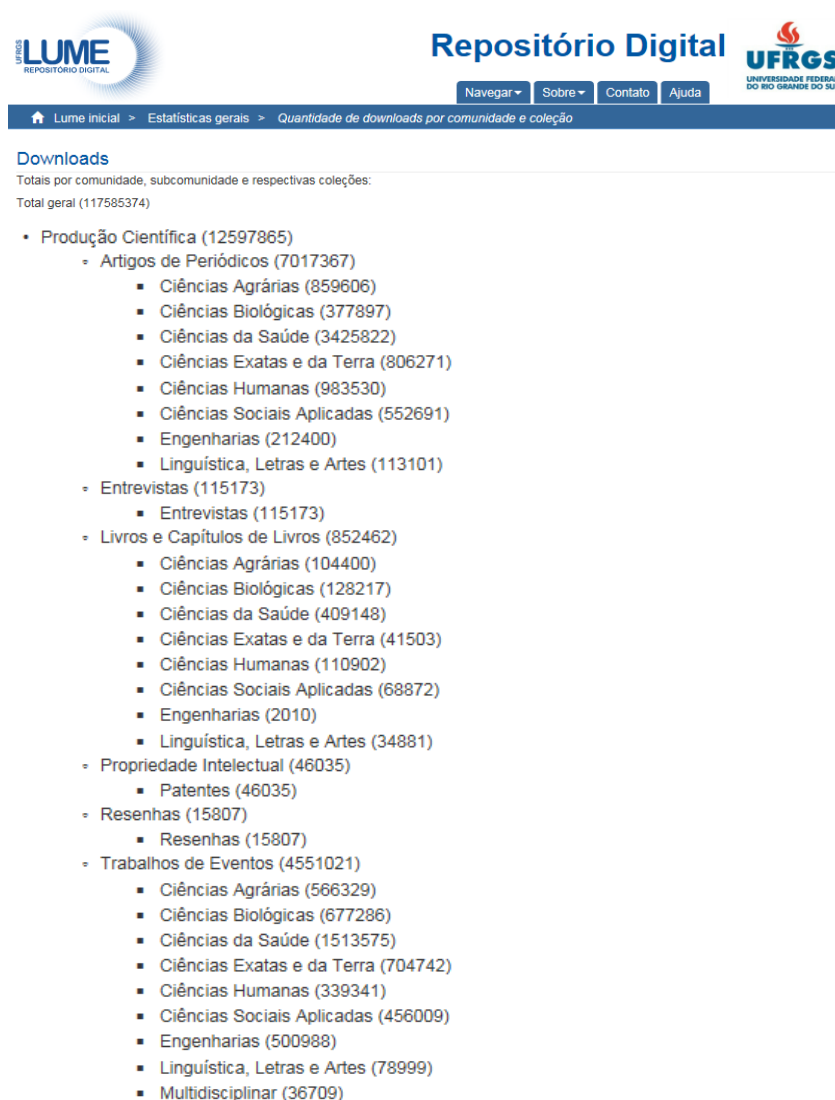


Figura 3: Número de *downloads*

b) itens com maior número de *downloads* (top 10)

A Figura 4 lista o título dos dez itens com maior número de *downloads* no segundo nível da hierarquia do Lume e inclui este dado entre colchetes.

Itens com mais downloads

Dez itens com maior número de downloads por comunidade ou subcomunidade:

- Produção Científica
 - Artigos de Periódicos
 - Medindo a ocorrência de doença : prevalência ou incidência? [129655]
 - Diagnóstico de enfermagem e intervenções em um paciente com falência de múltiplos órgãos : estudo de caso [69519]
 - Sistemas operacionais [67404]
 - A utilização da observação participante e da entrevista semi-estruturada na pesquisa de enfermagem [51144]
 - Medidas de associação em estudos epidemiológicos : risco relativo e odds ratio [50145]
 - Paralisia cerebral : novas perspectivas terapêuticas [43832]
 - Alfa-hidroxiácidos: aplicações cosméticas e dermatológicas [43593]
 - Velocidade de sedimentação globular (VSG) : informações úteis para o dia a dia [43375]
 - Nódulos de tireóide e câncer diferenciado de tireóide : consenso brasileiro [40958]
 - Livros e Capítulos de Livros
 - Avaliação e tratamento de feridas : orientações aos profissionais de saúde [277206]
 - Fisiologia humana : testes [106830]
 - Uso do leite para monitorar a nutrição e o metabolismo de vacas leiteiras. [27626]
 - Uso de provas de campo e laboratório em doenças metabólicas e ruminais de bovinos [9054]
 - Métodos de pesquisa [7869]
 - Aleitamento materno [6698]
 - Indicadores sanguíneos do metabolismo mineral em ruminantes. [6632]
 - Perfil metabólico em ruminantes : seu uso em nutrição e doenças nutricionais [6425]
 - Estratégias de prevenção de transmissão de germes multirresistentes : educação aos profissionais de saúde [6367]
 - Composição bioquímica do leite e hormônios da lactação. [6239]

Figura 4: Dez itens com maior número de *downloads*

c) número de acessos e *downloads*

Está disponível por meio de ícone próprio na página principal da comunidade, subcomunidade, coleção ou item.

Os dados são apresentados sob a forma de tabelas e gráficos, distribuídos anualmente (Figura 5). É possível visualizar, também, a distribuição mensal dos dados e restringir o período temporal, utilizando filtro de ano e mês. Ao final da página são relacionados, por *default*, os dez países com maior número de acessos e de *downloads*, com a opção de consulta à lista completa de países.

Estatísticas

Início da coleta: Jan. 2008

Teses e Dissertações defendidas na UFRGS

Estatísticas por ano por mês

Filtrar por data

Data inicial:

Ano

Todos

Mês

Todos

Data final:

Ano










Todos

Mês










Todos

Enviar

Downloads por país

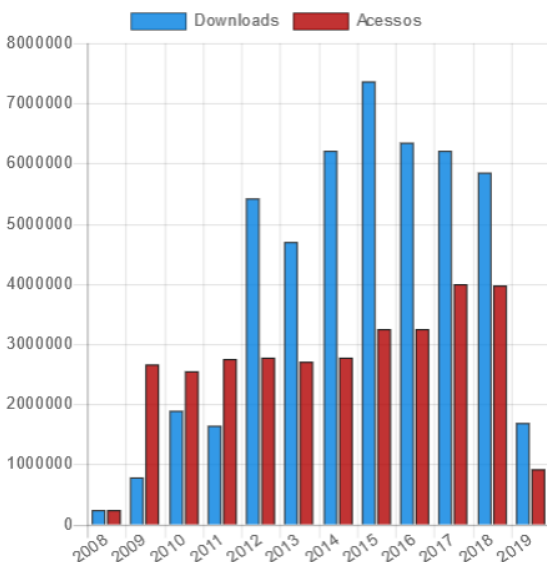
País	Downloads
 Brasil	34047534
 Alemanha	5812736
 Estados Unidos	2458414
 China	1648995
 Portugal	1385198
 Moçambique	471729
 UFRGS*	396855
? Outros	285105
 França	271665
 Angola	263901

Acessos por país

País	Acessos
 Brasil	23592250
 Estados Unidos	3504256
 China	1245240
 Portugal	919825
 Alemanha	602283
 UFRGS*	522717
 França	502035
? Outros	151591
 Japão	148849
 Moçambique	141890

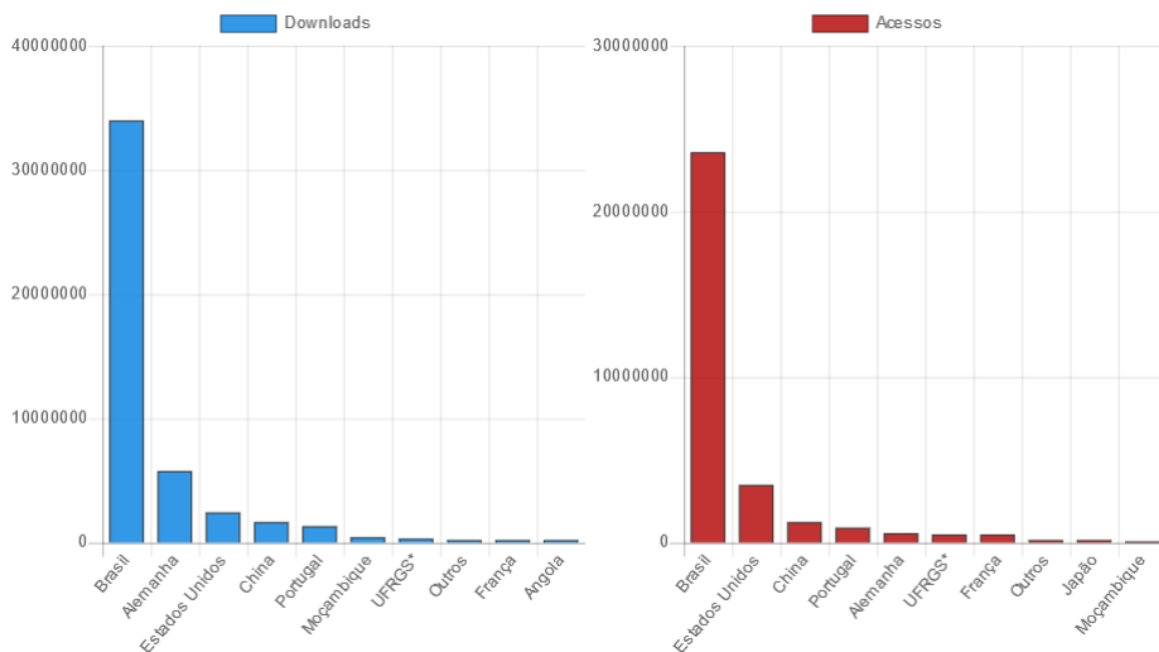
Estatísticas por ano

Ano	Downloads	Acessos
2008	231915	233535
2009	784272	2668554
2010	1883464	2543790
2011	1633944	2745703
2012	5421370	2773181
2013	4701780	2715252
2014	6217634	2773179
2015	7361810	3243374
2016	6351957	3249648
2017	6219448	4000839
2018	5845246	3982978
2019	1683438	928675
Total	48336278	31858708



Ano	Downloads	Acessos
2008	231,915	233,535
2009	784,272	2,668,554
2010	1,883,464	2,543,790
2011	1,633,944	2,745,703
2012	5,421,370	2,773,181
2013	4,701,780	2,715,252
2014	6,217,634	2,773,179
2015	7,361,810	3,243,374
2016	6,351,957	3,249,648
2017	6,219,448	4,000,839
2018	5,845,246	3,982,978
2019	1,683,438	928,675

160



*Os dados referentes a UFRGS estão incluídos nos dados do Brasil.

[Ver todos os países](#)

Figura 5: Número de acessos e *downloads* na subcomunidade de Teses e Dissertações defendidas na UFRGS

d) itens de um determinado autor ou assunto

Os dados são gerados a partir dos itens recuperados nos índices de autor e de assunto. Após a realização da busca, o link do ícone próprio das estatísticas fica no topo da lista de resultados (Figura 6).

Ao clicar no ícone, são listados, inicialmente, os cinco itens com maior e os com menor número de *downloads* do total de determinado autor depositados no Lume ou de assunto, seguidos do número de *downloads* e da data de entrada do item no repositório, conforme ilustra a Figura 7.

Na sequência, é exibida a distribuição mensal dos dados, com possibilidade de restrição do período temporal, e são apresentados, por *default*, os dez países com maior número de acessos e de *downloads*, além de *link* para a lista completa de países.



Figura 6: Itens para um autor

Estatísticas

Início da coleta: Jan. 2008

Estatísticas por autor "Costa, Janise Silva Borges da"

Itens da consulta com mais downloads

- [Anais das sessões temáticas e pôsters](#)
[Downloads:4955] [Data de entrada:16/12/2014]
- [Catalogação retrospectiva de livros nas bibliotecas da Universidade Federal do Rio Grande do Sul](#)
[Downloads:2114] [Data de entrada:6/6/2007]
- [Um modelo de integração entre sistemas de informação na Universidade Federal do Rio Grande do Sul : eventos e repositório digital](#)
[Downloads:1154] [Data de entrada:14/9/2012]
- [Integração entre bibliotecários e profissionais de tecnologia da informação da Universidade Federal do Rio Grande do Sul](#)
[Downloads:1013] [Data de entrada:16/11/2012]
- [O Processo de migração de sistema de automação de bibliotecas na Universidade Federal do Rio Grande do Sul, Brasil](#)
[Downloads:981] [Data de entrada:16/4/2010]

Itens da consulta com menos downloads

- [Artigos de periódicos em acesso aberto : citações distribuídas em repositórios institucionais](#)
[Downloads:20] [Data de entrada:15/1/2019]
- [LUME: MAIS VISIBILIDADE PARA A PRODUÇÃO CIENTÍFICA E ACADÊMICA DA UFRGS](#)
[Downloads:28] [Data de entrada:12/4/2017]
- [O papel dos repositórios institucionais como fonte de indicadores da comunicação científica](#)
[Downloads:33] [Data de entrada:21/12/2016]
- [Análise dos trabalhos apresentados nos seis anos da conferência BIREDIAL-ISTEC](#)
[Downloads:35] [Data de entrada:21/10/2017]
- [An implementation of technical revision in DSpace allowing open educational resource browser access](#)
[Downloads:64] [Data de entrada:15/12/2015]

Figura 7: Itens com maior e menor número de *downloads* para um autor

Considerações finais

O Lume - Módulo de estatísticas 2.0, agrega valor aos repositórios que utilizam o DSpace 5.8, ao possibilitar a recuperação e visualização gráfica de dados de acessos e de *downloads* nos vários níveis hierárquicos e, sobretudo, aos itens relacionados a um mesmo autor ou assunto.

Com a consulta aos dados já registrados pelo DSpace 5.8 no SOLR, foi possível melhorar o desempenho na recuperação e geração das estatísticas, bem como manter aquelas registradas originalmente pelo DSpace. Assim sendo, caso sejam criadas novas possibilidades de estatísticas no futuro, a integração será facilitada, pois o Módulo consulta as estatísticas padrão armazenadas pela ferramenta.

Referências

DIRECTORY OF OPEN ACCESS REPOSITORIES. OpenDOAR Statistics. 2019.

Acesso em: 9 maio 2019. Disponível em:

<http://v2.sherpa.ac.uk/view/repository_visualisations/1.html>.

LEITE, F. C. L. **Como gerenciar e ampliar a visibilidade da informação científica brasileira**: repositórios institucionais de acesso aberto. Brasília: IBICT, 2009. 124 p. Disponível em: <<http://livroaberto.ibict.br/handle/1/775>>. Acesso em: 18 maio 2019.

Integración de revistas desde repositorios institucionales hacia Open Journal Systems

Gonzalo Luján Villarreal¹, Ezequiel Manzur², Dolores García³, Marisa Raquel De Giusti⁴

¹ Dr. en Cs. Informáticas; Docente-Investigador UNLP; Subdirector CESGI, CIC.
Correo: gonzalo@prebi.unlp.edu.ar

² Analista Programador Universitario; Pasante PREBI-SEDICI, UNLP.
Correo: ezemanzur@sedici.unlp.edu.ar

³ Lic. en Comunicación Social; pasante CESGI, CIC.
Correo: dolores.garcia@sedici.unlp.edu.ar

⁴ Dr. en Cs. Informáticas; Investigador CIC; Directora PREBI-SEDICI UNLP; Directora CESGI, CIC. Correo: marisa.degiusti@sedici.unlp.edu.ar

Resumen

Muchos repositorios institucionales alojan colecciones de revistas, que o bien nacieron en formato digital, o bien fueron digitalizadas para aprovechar la plataforma de preservación y difusión que naturalmente brindan los repositorios. Usualmente, las revistas envían su producción a los repositorios, pero existen situaciones en las que las revistas requieren recuperar desde los repositorios los artículos y sus metadatos, como ser, cuando una revista migra hacia el mundo digital y comienza a utilizar un sistema de gestión editorial. Esto sucede a menudo en la Universidad Nacional de La Plata, que cuenta con un repositorio institucional (SEDICI), basado en *DSpace*, que aloja recursos de toda la Universidad, y un portal de revistas que funciona sobre *OJS* y sirve como espacio para revistas de toda la institución. Dado que muchas revistas ya han migrado hacia dicho portal o lo están haciendo, se implementó un mecanismo para integrar los recursos del repositorio en el portal de revistas. Se desarrolló un software capaz de mapear metadatos exportados desde *DSpace* hacia la estructura de *OJS*, y a partir de allí generar archivos XML para importar tanto metadatos como documentos completos. Se generó una interfaz web para que usuarios no informáticos puedan generar los XML, y se estableció un flujo de trabajo con editores para organizar y agilizar la importación artículos. Esta herramienta ha sido utilizada por 6 revistas, que han servido para ajustar algoritmos, optimizar el flujo de trabajo y mejorar la interfaz para hacerla más simple y brindarle mayor robustez al proceso.

Palabras clave: OJS, DSpace, Interoperabilidad, revistas

Abstract

Many institutional repositories host journals collections, either born digital or digitalized to take advantage of the preservation and dissemination platform offered by repositories. Usually, journals send their production to repositories, but there are situations in which journals need to retrieve articles and metadata from the repositories, e.g. a journal migrating to the digital world making use of an editorial management system. Those situations are often seen at Universidad Nacional de La Plata because of its institutional repository (SEDICI), based on DSpace, which hosts resources from the entire University, and a journals portal which works on OJS and serves as a space for journals of the entire University. Since many journals have already migrated to this portal, and many others are in the process, a mechanism to integrate repository's resources into the journals portal was implemented. A software capable of mapping metadata exported from DSpace to the OJS structure was developed, which is able to generate XML data to import both metadata and complete documents. A web interface was generated so users with none to zero technical knowledge can generate XML files, and a workflow with editors to organize and expedite the importation of articles was established. This tool has been used by 6 journals, which have served to adjust algorithms, optimize the workflow, make a simpler interface and give the process greater sturdiness.

Keywords: OJS, DSpace, Interoperability, journals

Introducción

Un repositorio institucional es un espacio virtual en el que se depositan documentos digitales, y cuyo propósito es gestionar, organizar, almacenar, preservar y difundir la producción resultante de la actividades de una organización. (De Giusti, 2014, p.12). Los repositorios institucionales suelen brindar distintos mecanismos de interoperabilidad para exponer sus recursos y así permitir que puedan ser consumidos por otros sistemas: servicios de análisis de citas, sistemas de gestión de recursos de investigación (CRIS), agregadores de recursos, entre otros. En un escenario típico, los repositorios reciben objetos digitales directamente de los productores (autores de artículos, tesis, editores, etc.) o por medio de intermediarios (desde la biblioteca, la oficina de títulos, la secretaría de investigación, etc), y sobre cada objeto recibido realizan los procesos correspondientes: control de calidad, verificación de licencias, catalogación, transformación, entre otros. Sin embargo, existen algunas situaciones en las que este escenario tradicional se invierte, y es el repositorio quien envía hacia otros sistemas los objetos digitales, para luego ser integrados en los circuitos de gestión. Un caso concreto se da con los repositorios institucionales que contienen revistas científicas, y que quieren comenzar a utilizar un sistema de gestión del ciclo editorial como *Open Journals System* (OJS).

Esta situación ha ocurrido en reiteradas ocasiones en La Universidad de La Plata, entre su repositorio institucional SEDICI (*Dspace*) y el Portal de Revistas (OJS 3). Muchas revistas que existen desde hace años (en algunos casos décadas) en el repositorio desean sumarse al Portal de Revistas, y para ello requieren algún mecanismo para importar todos los artículos y números a OJS. El software *Dspace* cuenta con varios mecanismos para exportar metadatos, como los archivos CSV y XML, y cuenta también con mecanismos para exponer metadatos, por ejemplo *OpenSearch* (Villarreal et al., 2017), OAI PMH o incluso campos META en los encabezados HTML. Asimismo, al momento de incorporar artículos ya publicados al sistema OJS, se puede hacer de forma manual (por ejemplo utilizando el plugin *QuickSubmit* de OJS), a través de comandos INSERT SQL o se puede utilizar el plugin de importación masiva de OJS, que utiliza el formato XML nativo para encapsular tanto los metadatos como los artículos en sus diferentes formatos (PDF, HTML, etc.).

El objetivo de este trabajo es presentar una herramienta que permite automatizar en gran parte el proceso de importación de revistas desde el repositorio hacia un portal de revistas en OJS a través del formato XML nativo, y proponer un flujo de trabajo para que los administradores de estos sistemas puedan realizar esta tarea de manera organizada y segmentada.

Materiales y metodología

Para realizar este trabajo se estudiaron los formatos mencionados para la exportación de metadatos en *Dspace* y los formatos de importación disponibles en OJS. Para realizar la exportación de metadatos se escogió el formato CSV por varios motivos: su simpleza a la hora de ser procesado (la mayoría de los lenguajes de programación provee soporte para interpretar archivos en formato CSV), porque ofrece la información mínima necesaria de las revistas para incorporarse a OJS, y porque permite agrupar en un único archivo todos los metadatos de todos números de una revista, lo que evita realizar múltiples exportaciones desde *Dspace*. Para la importación dentro de OJS, se seleccionó el formato XML nativo que utiliza el plugin de importación de artículos de OJS; si bien se evaluó la generación de consultas SQL personalizadas, este proceso puede ser muy complejo y propenso a errores (mientras que el plugin de importación funciona desde años y está en una etapa muy estable de su desarrollo). También se consideró la posibilidad de realizar la carga de artículos en OJS de forma manual, pero es evidente que este proceso requiere mucho tiempo por parte de los gestores y editores de las revistas, y es también propenso a errores originados a partir de fallas humanas. El uso de este plugin de importación es simple, el procesamiento de los mismos se realiza de manera muy eficiente, y se encuentra disponible en casi todas las instalaciones de OJS.

Una vez seleccionados los formatos de exportación e importación, se hizo necesario idear un método repetible y fácil de controlar para realizar la importación dentro de OJS. Esto sucede porque, por lo general, las importaciones implican

decenas de números y cientos de artículos, lo que requiere un trabajo metódico y cuidadoso para no pasar por alto ningún artículo, no cargar artículos repetidos y no mezclar metadatos entre artículos. En líneas generales, este flujo de trabajo consiste en una primera etapa de análisis del CSV exportado desde DSpace, segmentación en múltiples CSV de menor tamaño (uno por cada número de la revista) y generación de los documentos XML, y una segunda etapa en la cual se incorporan iterativamente los XML generados en la etapa anterior dentro de una instalación de OJS.

Procesamiento del CSV

En la etapa de procesamiento del CSV, se busca generar un mapeo al formato XML nativo capaz de ser importado en OJS. Para definir este mapeo se revisaron los metadatos de los artículos de revistas generados desde el repositorio, se relevaron los distintos esquemas disponibles en la exportación (típicamente, *Dublin Core*, aunque en ocasiones aparecen otros esquemas) y se analizaron requerimientos de normalización, mapeos y post-procesamiento de los metadatos. Además, se analizaron elementos disponibles en la exportación CSV que podrían aprovecharse para generar el XML nativo de OJS; por ejemplo, el metadato **dc.title[es]** de *Dublin Core* exportado en el CSV, que hace referencia al título en español del artículo, se presenta en el XML como el nodo `title` con el atributo de localización (*locale*) correspondiente. En este ejemplo, el mapeo en formato XML quedaría de la siguiente forma:

```
<title locale="es_ES"> Título en español </title>
```

De este modo, si el repositorio exporta metadatos en múltiples idiomas, al realizar el mapeo hacia XML se incorpora esta información, permitiendo así la incorporación de metadatos internacionalizados en OJS.

Al momento de realizar los mapeos, se observó que el repositorio SEDICI no exporta metadatos referidos a la sección de la revista a la que pertenece cada artículo. Sin embargo, este metadato es obligatorio en OJS: todos los artículos deben pertenecer a alguna sección en este sistema. Como solución, en todos los casos se mapean los artículos hacia una sección especial (llamada por lo general Importados o Imported), que se utiliza exclusivamente durante la etapa de migración. Esta sección es configurada desde OJS con varias limitaciones: los autores no pueden enviar artículos allí, no aparece en el índice de la revista y no se revisa por pares. De este modo, funciona como una sección interna, y su objetivo es alojar de manera temporal a los artículos recién importados, hasta tanto un editor los asigne a la sección correspondiente (reseñas de libros, artículos originales, resúmenes de tesis, etc.).

El formato XML nativo de OJS permite la incorporación de los artículos a texto completo como un elemento más del XML. Para ello, los archivos deben codificarse en base 64 (*IETF*, s/f) antes de ser adicionados al XML. De este modo,

al importar un XML a OJS, es posible importar tanto metadatos como documentos completos en formatos PDF, HTML, EPUB, etc. Sin embargo, los archivos CSV exportados por DSpace sólo incluyen metadatos, y no contienen ni los objetos digitales en sí ni la URL de acceso a los mismos. Afortunadamente, dicha URL es posible de obtener a partir del documento HTML referenciado por el metadato URI de cada recurso (en el caso de SEDICI, dicho metadato será **sedici.identifier.uri**). Entonces, la aplicación aquí presentada, además de realizar el mapeo de metadatos, utiliza el metadato URI para recuperar la página web (documento HTML) de cada recurso en el repositorio, luego analiza dicho HTML con el objetivo de recuperar el enlace al documento PDF, alojado en el nodo **META citation_pdf_url** (ver Imagen 1). Una vez obtenido dicho enlace, esta aplicación descarga desde el repositorio el documento PDF a través de una conexión HTTP y, una vez finalizada la descarga, codifica el documento PDF obtenido y lo incorpora dentro del XML que está siendo generado.

```

Institutional Websites" name="citation_title" />
<meta content="en" name="citation_language" />
<meta content="Villarreal, Gonzalo Luján" name="citation_author" />
<meta content="Salamone Lacunza, Paula" name="citation_author" />
<meta content="Vila, María Marta" name="citation_author" />
<meta content="De Giusti, Marisa Raquel" name="citation_author" />
<meta content="Manzur, Ezequiel" name="citation_author" />
<meta
content="http://sedici.unlp.edu.ar/bitstream/handle/10915/60507/Presentación
diapositivas.pdf?sequence=4" name="citation_pdf_url" />
<meta content="http://hdl.handle.net/10915/60507"
name="citation_abstract_html_url" />
<meta content="Open Repositories 2017 (Queensland, Australia, 2017)"
name="citation_conference_title" />
<meta content="2017-06-07T17:39:03Z" name="citation_online_date" />
</head><!--[if lt IE 7 ]> <body class="ie6"> <![endif]-->
<!--[if IE 7 ]> <body class="ie7"> <![endif]-->
<!--[if IE 8 ]> <body class="ie8"> <![endif]-->

```

Imagen 1: Representación del documento HTML de un artículo.

Importación en OJS

Los documentos XML obtenidos luego del mapeo deben ser importados uno a uno dentro de OJS. Este trabajo lo debe realizar un usuario con rol de editor o gestor, y luego de importar cada XML debe realizar una revisión manual de todos los artículos importados para:

- corregir posibles errores de mapeo, o completar datos faltantes (datos que el repositorio no poseía)
- ubicar los artículos en las secciones correspondientes
- publicar los números una vez revisados todos los artículos.

Cabe destacar que los documentos XML generados ya poseen la información sobre los números dentro de OJS. Esto evita al usuario la tarea de crear manualmente cada número, y de asignar los artículos importados a los

números correspondientes.

El desarrollo del software que da soporte a todo este proceso se realizó en dos etapas. En la primera, se creó una aplicación que funcionaba desde la línea de comandos, y cuyo funcionamiento era muy sencillo: dado un archivo CSV con los metadatos de la revista que se ingresaba como parámetro, sumado a algunos parámetros adicionales necesarios para procesar dicho archivo (por ejemplo la cantidad de registros a procesar o la sección de destino de los artículos), se procesaba y dividía el archivo CSV en varios archivos CSV de menor tamaño según el criterio ya explicado, y luego se procesaba a cada CSV con el objetivo de generar los archivos en formato XML correspondientes a cada número de la revista.

```
ezequiel@ezequiel-Lenovo-V330-15IKB:~/Escritorio/prueba/dspace2ojs$ php src/bin/csv2xml.php
10915-836 authors_group=autores into_section=IMPORTADOS split_csv=no
FILE 10915-836.csv : Let's start-----
Fetching from http://sedici.unlp.edu.ar/bitstream/handle/10915/14149/Documento_completo.pdf?sequence=1
Fetching from http://sedici.unlp.edu.ar/bitstream/handle/10915/14150/Documento_completo.pdf?sequence=1
Fetching from http://sedici.unlp.edu.ar/bitstream/handle/10915/14151/Documento_completo.pdf?sequence=1
Fetching from http://sedici.unlp.edu.ar/bitstream/handle/10915/14152/Documento_completo.pdf?sequence=1
Fetching from http://sedici.unlp.edu.ar/bitstream/handle/10915/14153/Documento_completo.pdf?sequence=1
```

Imagen 2: Generación de archivos XML a partir de un único archivo CSV (10915-836.csv), utilizando la aplicación desde la línea de comandos.

Como se mencionó previamente, la importación de una revista completa puede implicar cientos de artículos agrupados en decenas de números y volúmenes. Para organizar toda esta información, se agruparon los artículos en documentos XML por número. Esto fue posible gracias a la organización de recursos dentro del repositorio SEDICI, en el que cada revista posee una comunidad (Imagen 3), y cada número de las revistas es una subcomunidad dentro de la comunidad de la revista. La estructura de colecciones y comunidades de este repositorio es mucho más amplia, e incluye a las Unidades Académicas y colecciones especiales (*SEDICI*, s/f), pero esta estructura se gestiona de manera independiente a la jerarquía de comunidades de revistas, lo que permite asegurar que la información obtenida a partir de la jerarquía de revistas y números sólo contendrá elementos (artículos) de revistas.



Revistas

Explorar:

[Listar todo](#) [Autores](#) [Temas](#)

Subcomunidades en esta comunidad

47 al Fondo

Abreletras-Psicodiagnóstico

Acta Farmacéutica Bonaerense

Actas de Periodismo y Comunicación

Aletheia

Alp

Analecta Veterinaria

Anales de la Facultad de Ciencias Jurídicas y Sociales

Imagen 3: Subcomunidades de revistas en el repositorio SEDICI.

La información de las comunidades y subcomunidades se aprovechó también para generar los nombres de los archivos CSV: el CSV exportado por DSpace tiene por nombre el *handle* de la comunidad de la revista dentro del repositorio, y los archivos CSV y XML que representan un número específico toman por nombre el *handle* de la colección a la que pertenece cada uno, del CSV importado. Por ejemplo, a partir del *handle* de la revista *TE & ET* (hdl.handle.net/10915/836) el nombre del archivo exportado desde el repositorio resulta **10915-836.csv**, siendo la primera serie de números el prefijo del *handle* de SEDICI y la segunda, el identificador de la comunidad. Al momento de procesar este archivo, se obtiene el *handle* de cada número de la revista para nombrar a cada uno de los CSV derivados, manteniendo la estructura de prefijo de SEDICI y el identificador de colección que corresponde a cada número. El XML que se genera para cada número conserva el mismo nombre del CSV a partir del cual es generado, pero con la extensión xml. Entonces, a partir del ejemplo anterior, algunos de los CSV que representan a los números de la revista se pueden

nombrar como **10915_837.csv**, **10915_838.csv**, **10915_839.csv**, etc. Y el nombre de cada XML **10915_837.csv.xml**, **10915_838.csv.xml**, **10915_839.csv.xml**.

En la segunda etapa del desarrollo de esta aplicación se incorporó una interfaz de usuario web, que facilita el uso de la herramienta a personas sin conocimientos informáticos. Esta interfaz permite cargar el archivo CSV generado desde el repositorio y, una vez procesado en múltiples archivos CSV, permite generar uno a uno los XML correspondientes a cada CSV. Durante esta segunda etapa del desarrollo se agregaron también algunas funcionalidades nuevas, como la posibilidad de crear secciones de la revista para importar artículos, o generar automáticamente los números en OJS a partir de los metadatos del XML. Además, esta herramienta permite al usuario mantener un control sobre el proceso de migración, no solo mediante la generación a demanda de los documentos XML, sino también marcando (archivando) los archivos XML ya importados, o incluso brindando la posibilidad de procesar todos los archivos CSV juntos (muy útil en el caso de revistas de menor tamaño).

10915-836	Xml	Csv	Process	Archived
2	10915_838.csv	10915_838.csv.xml		<input type="button" value="Archived"/>
3	10915_839.csv	-	<input type="button" value="Process"/>	<input type="button" value="Archived"/>
4	10915_840.csv	-	<input type="button" value="Process"/>	<input type="button" value="Archived"/>
5	10915_841.csv	-	<input type="button" value="Process"/>	<input type="button" value="Archived"/>
6	10915_842.csv	-	<input type="button" value="Process"/>	<input type="button" value="Archived"/>
7	10915_18268.csv	-	<input type="button" value="Process"/>	<input type="button" value="Archived"/>
8	10915_25508.csv	-	<input type="button" value="Process"/>	<input type="button" value="Archived"/>

Imagen 4: Interfaz de usuario del sistema.

Decisiones particulares sobre los mapeos de metadatos

Los dos sistemas con los que se trabaja aquí, OJS y DSpace, utilizan estructuras de metadatos muy diferentes; sumado a esto, las herramientas de importación y exportación utilizadas realizan sus propias adaptaciones sobre dichos metadatos: algunos metadatos se combinan en uno solo al ser exportados, otros metadatos no existen o no son exportados, etc. Asimismo, la estructura jerárquica del formato XML nativo de OJS, si bien es muy sencilla (son relativamente pocos elementos), posee algunas restricciones en cuanto a los valores posibles y en cuanto a la obligatoriedad de muchos de sus componentes. A continuación se presenta una vista simplificada del formato XML nativo de OJS, donde puede apreciarse la estructura jerárquica del mismo y los elementos que

considera:

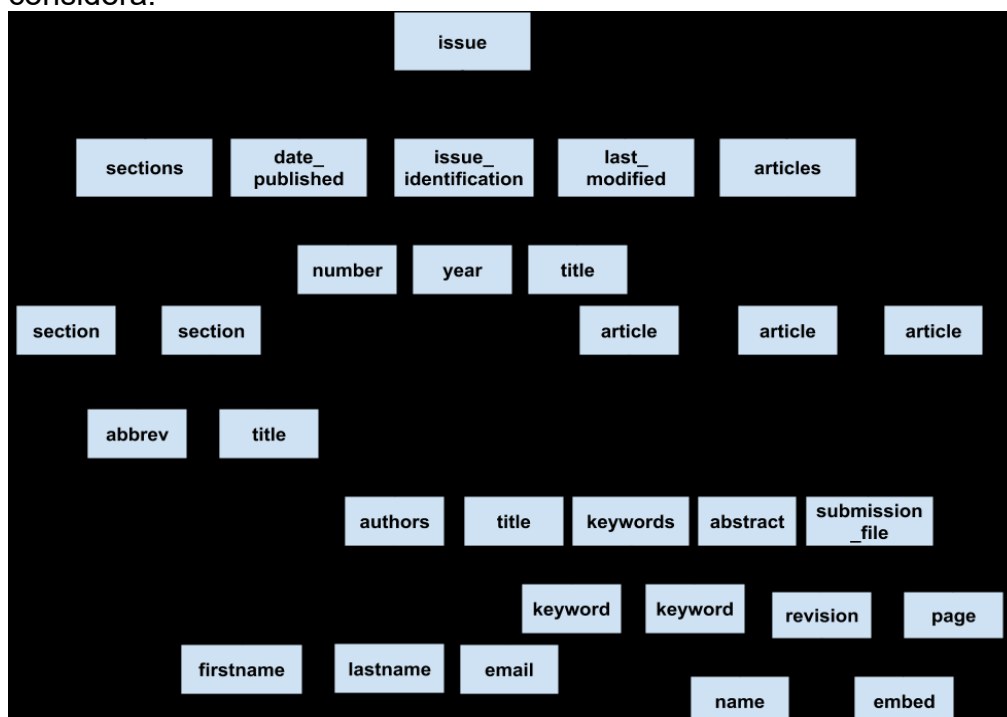


Imagen 5: Estructura jerárquica del XML nativo.

A diferencia del formato XML nativo, las exportaciones en CSV generadas por DSpace se componen de decenas de columnas con metadatos, que pueden variar según cada tipo de documento (artículo, tesis, libro, etc.) y que pueden pertenecer a más de un esquema de metadatos, como ser Dublin Core, Mods, elementos propios de DSpace o esquema de metadatos propio del repositorio. La Tabla 1 expone los mapeos realizados entre metadatos de ambos formatos.

Nodo XML		Metadato en el CSV	Descripción
issue_identification	number	-	El software le asigna un orden creciente a los números; pero puede no coincidir con el real. Este metadato debe revisarse y modificarse, de ser necesario.
	year	dc.date.issued	
	title	-	Se asigna un título relacionado con el número. Este metadato se debe corregir o eliminar.

date_published		dc.date.issued	Este metadato lo exige el XML nativo, por eso se incorpora el mismo dato year
last_modified			
section	abbrev	-	Este metadato lo exige el XML nativo, por eso se le asigna el mismo valor que el título.
	title	-	Nombre de la sección donde serán importados los artículos. Se ingresa desde el software.
article	title	dc.title	Metadato que se puede discriminar por idioma.
	abstract	dc.description.abstract	Metadato que se puede discriminar por idioma.
	keyword	sedici.subject.keyword	Las palabras clave se importan separadas por “;” y el software se ocupa de generar un nodo para cada una. Este metadato que se puede discriminar por idioma.
	author	sedici.creator.person	Los autores vienen separados por “ ”, el software lo procesa y obtiene nombre y apellido para cada autor
	email	-	Este metadato no es exportado desde SEDICI y es un campo obligatorio en el XML nativo, por lo que se crea un mail falso, y se corrige en OJS.
	revision	name	Nombre del formato del archivo (PDF)
		embed	De este metadato se obtiene la URI al documento. El software descarga el documento, lo codifica en base 64 y lo incorpora en el XML.
	pages		

Tabla 1: Mapeos de metadatos entre el esquema XML nativo de OJS y los metadatos exportados desde DSpace.

Resultados y trabajos futuros

Este desarrollo fue utilizado con éxito en la migración de 6 revistas desde el repositorio SEDICI hacia Portal de Revistas de la Universidad Nacional de La Plata:

- Journal of Computer Science and Technology (JCST)
- *Tecnología en Educación y Educación en Tecnología (TE&ET)*
- *Económica*
- *Analecta Veterinaria*
- *Biología Acuática*
- *Anales JURSOC*

Es importante destacar que algunas de estas revistas (JCST y TE&ET) funcionan sobre un OJS externo, gestionado por la Facultad de Informática de la UNLP. El desarrollo aquí propuesto no está atado al OJS del Portal de Revistas de la UNLP, sino que genera documentos XML compatibles con cualquier sistema OJS (versión 3.0 o superior).

En líneas generales, se estima que en aproximadamente 10 o 15 minutos un editor puede importar, revisar y publicar un número completo de entre 10 y 12 artículos. Los tiempos para la generación del XML a importar son más variables, ya que entra en juego la velocidad de conexión con el repositorio y el tamaño de los archivos que se descargan. De todos modos, en las migraciones realizadas hasta ahora, los tiempos de generación de los XML oscilaron entre unos pocos segundos hasta unos 4 minutos para revistas cuyos números poseen muchos artículos y archivos PDF de gran tamaño. Por lo tanto, si se suman los tiempos de generación del XML, descarga desde la herramienta, carga y procesamiento en OJS, revisión y corrección de metadatos, y finalmente publicación del número, el tiempo total no supera los 20 minutos por número. Es evidente que la alternativa manual a este proceso, en la que el usuario debe descargar a mano cada PDF desde el repositorio y además copiar uno a uno los metadatos de cada artículo desde el repositorio hacia OJS, requerirá mucho más tiempo y será más propenso a errores y falla. Si bien el proceso para migrar una revista no está automatizado por completo (se debe importar los XML de a uno, revisar y completar los metadatos de los artículos y luego publicarlos), tanto el uso de la herramienta, como el flujo de trabajo propuesto reducen, de forma considerable, el tiempo para migrar revistas hacia el portal.

Otra ventaja de esta herramienta es el soporte que brinda a la hora de organizar el flujo de trabajo, en particular la segunda etapa en la que la misma aplicación genera una sección especial donde se importan los artículos -para evitar confusiones con los envíos en proceso editorial-, crea los números y permite gestionar el progreso de importación a través de la interfaz de usuario.

Más allá de las ventajas del uso de esta herramienta, a medida que más usuarios la utilizaron comenzaron a surgir algunas situaciones en las que se planteaba la necesidad de incorporar o de mejorar funcionalidades. En primera

medida, se observó la necesidad de considerar otros esquemas de metadatos más allá de Dublin Core; cabe recordar que, como se observa en la tabla 1, esta aplicación utiliza de momento sólo metadatos *Dublin Core* y del esquema propio de SEDICI. Si bien el diseño de la aplicación considera los mapeos entre metadatos de manera especial, ya que considera una sección interna de configuración de mapeos y transformaciones de metadatos, esto no ha sido puesto a prueba con otros repositorios y tampoco es evidente para los usuarios finales. El problema puntual se manifiesta cuando se requiere su uso desde otro repositorio, con sus propios metadatos, o cuando se deben incorporar metadatos que no existen, como puede ser el caso del correo electrónico del autor. Una posible solución para este último caso podría ser la incorporación de los metadatos de forma dinámica y configurable desde la aplicación, de modo que cada usuario pueda elegir con qué esquemas representar a cada uno del XML nativo.

En segunda medida, como se mencionó previamente, este desarrollo hace uso de la estructura jerárquica del repositorio SEDICI, valiéndose de las colecciones y comunidades que representan a las revistas y a sus números. Si bien esto resultó una gran ventaja en el contexto de la UNLP, esto no es posible de aplicar en repositorios que se encuentran estructurados de manera diferente. Por ejemplo, el repositorio CIC-Digital¹ utiliza las comunidades para agrupar recursos de los distintos centros de investigación, y carece de una comunidad o colección dedicada para las revistas. De todos modos, dado que los metadatos exportados por DSpace incluyen información sobre el número y el año de cada artículo, con algunos cambios en el análisis del CSV inicial podría realizarse la separación en múltiples CSV y a partir de allí generar los XML. Cabe destacar que esto requerirá también una etapa de filtrado, ya que en principio el CSV exportado podría incluir recursos de todo tipo, y no solo artículos de revistas propias.

Otra carencia de este desarrollo es la incorporación de artículos en más de un formato. Esto se debe al uso del elemento META **citation_pdf_url**, que es único por cada artículo, incluso si el artículo posee varias versiones (EPUB, HTML, XML JATS, etc.). Para resolver este problema, la herramienta podría obtener los enlaces a los documentos a partir del cuerpo del HTML del repositorio, en vez de utilizar el elemento META del encabezado.

Finalmente, sería deseable permitir al usuario cargar distintas fuentes de datos en vez de un archivo CSV exportado desde el repositorio. Resulta particularmente interesante la posibilidad de brindar soporte al formato XML utilizado por el protocolo OAI PHM. De este modo, el usuario podría especificar una interfaz de acceso OAI a la colección o comunidad dentro de un repositorio, que incluso podría seguir directrices de interoperabilidad estandarizadas (ej. *Driver* u *OpenAire*), cosechar desde allí los metadatos en formato XML, y tomar ese XML como insumo de entrada para generar el o los XML en formato nativo de OJS.

1 Consultar en <https://digital.cic.gba.gob.ar/>.

Referencia

1. Batch Metadata Editing, DuraSpace Wiki (en línea). <https://wiki.duraspace.org/display/DSDOC5x/Batch+Metadata+Editing>.
2. IETF (s/f). The Base16, Base32, and Base64 Data Encodings (en línea). <https://tools.ietf.org/html/rfc4648>.
3. De Giusti, M. R. (2014). Una metodología de evaluación de repositorios digitales para asegurar la preservación en el tiempo y el acceso a los contenidos (Doctoral dissertation, Facultad de Informática).
4. De Giusti, M., García, D., Manzur, E., Villarreal, G. L., & Folegatto, L. E. (2018). Importación de artículos en OJS desde Dspace. Recuperado de <http://repositorio.pucp.edu.pe/index/handle/123456789/133195>.
5. DSpace importing and exporting items via Simple Archive Format (en línea). <https://wiki.duraspace.org/display/DSDOC5x/Importing+and+Exporting+Items+via+Simple+Archive+Format#ImportingandExportingItemsviaSimpleArchiveFormat-ExportingItems>.
6. DSpace2OJS, espacio Github SEDICI UNLP (en línea). <https://github.com/sedici/dspace2ojs>.
7. OJS Data Import and Export (en línea). <https://docs.pkp.sfu.ca/admin-guide/en/data-import-and-export>.
8. Open Journal Systems and Dataverse Integration – Helping Journals to Upgrade Data Publication for Reusable Research (en línea). https://openscholar.mit.edu/sites/default/files/dept/files/the_code4lib_journal_-_open_journal_sys.data_publication_for_reusable_research.pdf.
9. SEDICI (s/f). Preguntas frecuentes (en línea). <http://sedici.unlp.edu.ar/pages/FAQ#organizacion>.
10. Villarreal, G. L., Manzur, E., Vila, M. M., De Giusti, M. R. (2017). Interoperabilidad con repositorios digitales: uso de OpenSearch en sitios web institucionales (en línea). En *Conferencia Internacional sobre Bibliotecas y Repositorios Digitales de América Latina (BIREDIAL-ISTEC'17) y Simposio Internacional de Biblioteca Digitales (SIBD'17) (Vol. 7)*. Recuperado de <http://sedici.unlp.edu.ar/handle/10915/63566>.

A ciência e o livro eletrônico: o projeto BRAPCI Livros

Rene Faustino Gabriel Junior¹; Nanci Oddone²

¹ Prof. Dr. do Programa de Pós-Graduação em Ciência da Informação (PPGCIN) - Universidade Federal do Rio Grande do Sul (UFRGS). Email: renefgj@gmail.com

² Profa. Dra. Programa de Pós-Graduação em Biblioteconomia - Universidade Federal do Estado do Rio de Janeiro (UNIRIO). Email: neoddone@unirio.br

Resumo

Na comunicação científica os livros têm um papel fundamental na disseminação de conhecimentos científicos. Entretanto o mercado editorial está tendo que se adaptar para garantir a sustentabilidade da economia cultural, científica e educacional do país. Nesse contexto os *eBook* vem ganhando posição pela sua facilidade de acesso e distribuição. Este trabalho se encaixa no eixo de portal de periódicos e publicações eletrônicas do evento. O objetivo da pesquisa é estimular a visibilidade e assegurar o acesso aos livros acadêmicos publicados em acesso aberto no Brasil e no exterior, reunindo e preservando obras natos digitais, com ênfase nas áreas de Biblioteconomia, Ciência da informação, Comunicação, Arquivologia e Museologia. O projeto tem duas vertentes, uma tecnológica que subsidia o desenvolvimento de um aplicativo (software para web), e uma aplicada com a identificação de canais, plataformas e/ou instituições que disponibilizam obras técnicas e acadêmicas de autores brasileiros em acesso aberto e sua organização e disponibilização no aplicativo. O projeto encontra-se na fase de identificação e coleta de obras para indexação, e espera-se que o projeto Brapci Livros contribua para ampliar a visibilidade dos Livros produzidos na área, bem como garantir o acesso a preservação digital, respeitando os direitos autorais e patrimoniais.

Palavras-chave: eBooks. Preservação Digital. Acesso aberto.

Abstract

In scientific communication, books play a fundamental role in the dissemination of scientific knowledge. However, the publishing market is having to adapt to ensure the sustainability of the cultural, scientific and educational economy of the country. In this context the eBook is gaining position by its ease of access and distribution. This work fits into the portal axis of periodicals and electronic publications of the event. The objective of the research is to stimulate visibility and ensure access to academic books published in open access in Brazil and abroad, bringing together and preserving original digital works, with emphasis in the areas of Librarianship, Information Science, Communication, Archivology and Museology. The project has two strands, a technological one that subsidizes the development of an application (web software), and an applied one with the identification of channels, platforms and / or institutions that make available technical and academic works of Brazilian authors in open access and its organization and in-app provisioning. The project is at the stage of identification

and collection of works for indexing, and it is expected that the Brapci Books project will contribute to increase the visibility of the Books produced in the area, as well as guarantee access to digital preservation, respecting copyright and assets.

Keywords: eBooks. Digital Preservation. Open Access.

Introdução

Até recentemente, as bibliotecas universitárias se valiam da publicação regular dos catálogos de editoras comerciais e universitárias para identificar publicações que fossem do interesse de estudantes, professores e pesquisadores que frequentam essas bibliotecas. Os estudantes, profissionais e leitores em geral, por sua vez, se serviam das numerosas livrarias que, em cada cidade, exibiam os livros publicados pelas principais editoras nacionais em diferentes áreas do conhecimento. Nesse contexto, a cadeia produtiva do livro técnico e acadêmico era mantida por um modelo de negócio plenamente consolidado, que envolvia agentes tradicionais, com papéis especializados e bem definidos e métodos de trabalho próprios, que se articulavam para garantir a sustentabilidade dos circuitos científico e educacional do país.

Ao longo das últimas décadas, no entanto, a globalização da economia e da cultura, a popularidade das tecnologias eletrônicas, assim como a emergência das redes sociais e o progressivo fortalecimento dos valores e das práticas defendidas pelo Movimento de Acesso Aberto à Informação Científica, entre outros fatores pertinentes, transformaram a comunicação científica e a cadeia produtiva do livro, impondo mudanças significativas à produção, à circulação e ao consumo das publicações técnicas e acadêmicas. Atualmente existem diferentes modelos de negócio em funcionamento (SERRA, 2014), desde a impressão tradicional até a publicação através de plataformas digitais independentes, reconfigurando de modo irreversível o ciclo da produção editorial.

No contexto da comunicação científica contemporânea, centrada no ambiente digital e marcada pela multiplicidade de canais e veículos de transmissão e distribuição da produção intelectual das universidades e dos programas de pós-graduação brasileiros, observa-se que a baixa visibilidade e a dificuldade de acesso são dois dos aspectos que mais prejudicam a circulação e o consumo dos resultados da ciência nacional, inclusive entre os próprios pesquisadores. Considerando que as livrarias físicas e os eventos científicos presenciais estão deixando de ser espaços privilegiados para divulgar a produção editorial dos cientistas, propõe-se a criação do portal de publicações acadêmicas BRAPCI Livros, uma plataforma que objetiva contribuir para tornar a produção acadêmica e técnica da Ciência da Informação mais disponível para estudantes, docentes e profissionais das áreas do conhecimento relacionadas.

O projeto BRAPCI Livros

A ideia de reunir e disponibilizar os livros publicados em meio digital e acesso aberto pelos pesquisadores da Ciência da Informação surgiu de uma dupla

constatação: por um lado, verifica-se que a área mantém uma produção editorial regular e consistente; por outro observa-se a dispersão e a baixa visibilidade dessa mesma produção no espaço latino-americano. Tal descontinuidade entre produção e uso acaba muitas vezes silenciando vozes e minimizando interlocuções fundamentais para a consolidação do campo e para a preservação do patrimônio documental e bibliográfico da disciplina, historicamente negligenciado. Nesse contexto, o projeto visa ampliar a visibilidade e assegurar o acesso aos livros acadêmicos publicados em acesso aberto no Brasil e no exterior, reunindo e preservando obras nato-digitais, com ênfase nas áreas de Biblioteconomia, Ciência da informação, Comunicação, Arquivologia e Museologia.

Desenvolvido em parceria entre o Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Rio Grande do Sul (PPGCIN/UFRGS) e o Programa de Pós-Graduação em Biblioteconomia da Universidade Federal do Estado do Rio de Janeiro (PPGB/UNIRIO), o Projeto da BRAPCI Livros envolve ampla colaboração científica e tecnológica em todos os aspectos das atividades acadêmicas: ensino, pesquisa, extensão e inovação.

O projeto abrange várias questões a serem problematizadas para identificar as soluções mais adequadas ao sucesso da iniciativa. Na perspectiva tecnológica, por exemplo, já está em fase de testes o projeto piloto do aplicativo (software para web) que atende um Perfil de Aplicação (PA) alinhado ao modelo conceitual do FRBR, permitindo a representação, a organização, o armazenamento e o acesso aos *eBooks*. O pleno funcionamento do aplicativo enseja outros desafios a serem superados. Entre eles incluem-se a descoberta e o mapeamento dos recursos já dispersos pela web, abrangendo canais, plataformas e/ou instituições que tenham disponibilizado obras acadêmicas em acesso aberto nas áreas selecionadas; a obtenção das necessárias autorizações legais; a elaboração de políticas de seleção, formação e preservação de acervos digitais nas temáticas indicadas; a definição de rotinas operacionais, incluindo questões relacionadas a formatos, padrões e terminologias; e ainda a constituição de um conselho editorial, que legitime as ações e decisões a serem implementadas. A Figura 1 apresenta o piloto do site da !Brapci Livros!

Figura 1 – Página inicial da Brapci Livros



Fonte: <http://www.brapci.inf.br/books>

A questão dos conceitos e da nomenclatura

Existem diversas definições técnicas de eBook, que é uma abreviação do termo inglês *eletronic book* e significa livro em formato digital (GRAU; ODDONE; DOURADO, 2014). O eBook pode ser uma versão digital de uma obra impressa ou pode ter sido lançado apenas em formato eletrônico (COUTINHO; PESTANA, 2015). Quanto aos formatos, os mais conhecidos são o ePUB – formato ideal para *eReaders*, *Tablets* e *SmartPhones*, pois possibilita alterar o tipo e o tamanho da fonte do texto – e o *Portable Document Format* (PDF) – que possibilita a leitura estável em dispositivos eletrônicos e computadores. O PDF foi desenvolvido na década de 1980 como um formato de arquivo para representar documentos independentemente da aplicação do software, do hardware e do sistema operacional (COUTINHO; PESTANA, 2015). Hoje, no entanto, existem muitos outros formatos, alguns abertos e outros proprietários.

Embora ainda não se tenha chegado a um consenso em termos conceituais, a literatura periódica da área demonstra o crescente interesse pela temática dos eBooks, seu uso e seu impacto social e científico. Para dar conta da multiplicidade de novos enfoques e estudos, os autores estão realizando um levantamento para caracterizar tendências contemporâneas no âmbito dessa temática, tanto no Brasil como no exterior (GIMÉNEZ TOLEDO; CÓRDOBA RESTREPO, 2018).

As iniciativas pioneiras

Até o momento já foram identificadas algumas dezenas de iniciativas que envolvem livros acadêmicos disponíveis *online* em acesso aberto, tanto no Brasil como no exterior. A mais antiga de todas elas é sem dúvida o Projeto Gutenberg. Na Figura 2 está o Scielo Livros, por exemplo, que surgiu em 2012, foi pioneiro no Brasil (DOURADO; ODDONE, 2012). Hoje o país dispõe também da Coleção Cultura Acadêmica, mantida pela Editora Unesp, e dos portais Livro Aberto, do IBICT, e Livro Eletrônico, do CNPq, sem mencionar as várias editoras universitárias e comerciais que disponibilizam títulos de suas coleções em acesso aberto.

Figura 2 – Página principal do Scielo Livros



Fonte: <http://books.scielo.org/>

No exterior devem ser citadas a *OAPEN Library* (Figura 3) e o *Diretorio de Livros em Acesso Aberto* (DOAB) (Figura 4), iniciativas pioneiras que resultaram do esforço coletivo de algumas editoras universitárias dos Países Baixos.

Figura 3 - Site OAPEN Library



Fonte: <http://www.oapen.org/home>

Figura 4 - Diretório de Livros em Acesso Aberto (DOAB)



Fonte: <https://www.doabooks.org>

Iniciativas mais recentes são a *OpenEdition Books*, a *Open Textbook Library*, o Coletivo Radical OA, e o grupo *ScholarLed*, que reúne a *Open Humanities Press*, a *Open Book Publishers*, a *Mattering Press*, a *MayFly Books*, a *Meson Press*, a *Punctum Books* e a *re.press*. Além disso, editoras tradicionais como Springer, Cambridge, Westminster e Bristol mantêm coleções interdisciplinares em acesso aberto, disponíveis nos respectivos endereços.

Os portais temáticos

Em termos de portais de livros acadêmicos em acesso aberto para áreas do conhecimento específicas, como se pretende que seja a *!Brapci Livros!*, existe a Base de Física, apresentada na Figura 5, que oferece livros digitais de Ciências Exatas como Física, Astronomia, Matemática, Química, Biologia e Engenharia. A interface dessa base, porém, inclui poucos recursos e os livros são listados de acordo com seus títulos, em ordem alfabética. As obras são sempre referenciadas a partir de outras bases e plataformas, a exemplo do Projeto Gutenberg.

Figura 5 - Physics Database



Fonte: <http://physicsdatabase.com/>

Reflexões e expectativas

Quando se considera a dificuldade de obter o apoio de editoras comerciais ou universitárias e a inviabilidade de arcar com os altos custos de impressão dos livros, acredita-se que o *eBook* tem sido uma boa opção para os pesquisadores divulgarem os resultados de suas pesquisas. Entretanto tem-se observado que essa produção encontra-se pulverizada nas páginas pessoais dos pesquisadores, nos sites das editoras universitárias (BUFREM, 1992; DOURADO, 2012), ou ainda nos repositórios institucionais (COSTA; PAVÃO; HOROWITZ, 2014). Além de não atingir sua plenitude e seu potencial na comunicação científica, não alcançando o reconhecimento de seus pares, essa produção corre o risco de desaparecer na internet, tanto pela volatilidade das tecnologias como pela baixa sustentabilidade dos empreendimentos que disponibilizaram esses conteúdos, com o possível fechamento das editoras e a indisponibilidade de suas coleções.

Espera-se que o projeto !Brapi Livros! possa contribuir para consolidar a produção científica das áreas selecionadas, para ampliar a visibilidade dos livros, publicados pelos pesquisadores do campo, para garantir o acesso e a preservação digital desses conteúdos acadêmicos, e para fazer circular o conhecimento científico

especializado, respeitando os direitos autorais e patrimoniais, como hoje ocorre com a Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI) (BUFREM, et. al, 2010).

Referências

BUFREM, L. S. Editoras universitárias no Brasil: uma crítica para reformulação na prática. **Revista da Escola de Biblioteconomia da UFMG**, v. 21, n. 1, 1992.

BUFREM, L. S.; COSTA, F. D. O.; GABRIEL JUNIOR, R. F.; PINTO, J. S. P. Modelizando práticas para a socialização de informações: a construção de saberes no ensino superior. **Perspectivas em Ciência da Informação**, v. 15, n. 2, p. 22-41, 2010.

COSTA, J. S. B.; PAVÃO, C. M. G.; HOROWITZ, Z. Análise das ações implementadas para povoamento de repositórios institucionais no Brasil e em Portugal. **Cadernos BAD**, n. 2, p. 149-152, 2014.

COUTINHO, P.; PESTANA, O. Ebooks: evolução, características e novas problemáticas para o mercado editorial. **Páginas A&B, Arquivos e Bibliotecas**, n. 3, p. 169-195, 2015.

DOURADO, S. M. **Identificando a inovação editorial na cadeia produtiva do livro universitário brasileiro**. 110 f. 2012. Dissertação (Mestrado) – Instituto de Ciência da Informação, Universidade Federal da Bahia, Salvador, 2012.

DOURADO, S. M.; ODDONE, N. SciELO livros: inovação editorial para a comunicação em CT&I. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13. 2012, Rio de Janeiro. **Anais...** Rio de Janeiro, 2012.

GIMÉNEZ TOLEDO, E.; CÓRDOBA RESTREPO, J. F. (eds.). **Edición académica y difusión: libro abierto en Iberoamérica**. Bogotá: Universidad del Rosario; Comares, 2018.

GRAU, I.; ODDONE, N.; DOURADO, S. E-books, livros digitais ou livros eletrônicos?: um estudo terminológico. In: ENCONTRO NACIONAL DE



PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 15., 2014, Belo Horizonte. **Anais...**
Florianópolis: UFSC, 2013.

SERRA, J. G. **Livro digital e bibliotecas**. Rio de Janeiro: Fundação Getulio
Vargas, 2014.

Los datos de investigación en las Humanidades – periódicos decimonónicos

Isabel Galina Russell¹, Miriam Peña Pimentel², Ernesto Priani Saisó³

¹ Instituto de Investigaciones Bibliográficas, Universidad Nacional Autónoma de México (UNAM). Correo: igalina@unam.mx

² Instituto de Investigaciones Bibliográficas, Universidad Nacional Autónoma de México (UNAM). Correo: miriampp@unam.mx

³ Facultad de Filosofía y Letras, Universidad Nacional Autónoma de México (UNAM). Correo: epriani@filos.unam.mx

Resumen

Introducción: El OCR de periódicos digitalizados del siglo XIX pueden entenderse como "datos crudos" para la realización de investigación histórica hemerográfica y ofrecemos una reflexión en torno a cómo deben integrarse estos datos al repositorio para poder ser utilizado para estos fines. Este trabajo forma parte de los resultados de "Oceanic Exchanges: Tracing Global Information Networks in Historical Newspaper Repositories", un proyecto de minería de datos en repositorios de periódicos digitalizados. Materiales y metodología: OcEx está compuesto por 6 equipos de investigación en 9 países y busca modelar patrones de flujo de información en periódicos del XIX. México participa con el repositorio Hemeroteca Nacional Digital de México (HNDM) que contiene más de 7 millones de imágenes y sus correspondientes archivos en XML producto del proceso de OCR. A partir de utilizar este repositorio para realizar un proyecto de minería de datos humanístico hacemos una reflexión en cómo pueden ser concebidas las colecciones digitales y sus datos de tal forma que pueden ser utilizados como datos crudos para investigación hemerográfica. Resultados y conclusiones: El uso de repositorios para custodiar colecciones de datos crudos permite compartir y reutilizar esta información. Sin embargo, es necesario una descripción bibliográfica de la digitalización. Para lograr esto es necesario aproximarse a los datos resultantes de una digitalización, desde una perspectiva crítica y no únicamente tecnológica. Este trabajo representa un primer acercamiento y reflexión a cómo este tipo de datos deben de ser almacenados en repositorios para que sean adecuadamente interpretados y utilizados en la investigación.

Abstract

Introduction: OCR from digitised 19th century newspapers can be viewed as "raw data" for historical periodical researchers. In this talk we offer an approach on how this data can be integrated into a repository so that it can be useful for this type of work. This paper presents some of the results of "Oceanic Exchanges: Tracing Global Information Networks in Historical Newspaper Repositories (OcEx)", an international data mining project focused on repositories of historical digitized newspapers. Materials and

methodology: OcEx is a project consisting of 6 research teams in 9 different countries that seeks to model information flow in 19th century newspapers using data mining techniques. Mexico participates with the national newspaper library, the Hemeroteca Nacional Digital de México (HNDM), with over 7 million images and their corresponding XML OCR files. Using the data mining work we did on the HNDM as a case study, we reflect on how these digital collections and the data can be designed in order to be used as raw data for periodical studies. Results and conclusions: The use of repositories for raw data collection allows for sharing and reuse. However, a bibliographical description of the digitization is necessary. In order to do this a critical and analytical and not just technological approach to the digitization process is necessary. This paper is an initial approximation and reflection on how historical newspaper collections can be stored in repositories in such a way that they can be interpreted and used adequately for humanistic research.

Introducción

Las publicaciones formales académicas, tales como artículos de revistas o capítulos de libro, han sido de interés primordial para muchos repositorios y el movimiento de acceso abierto. Los repositorios sin embargo, pueden albergar una vasta cantidad de materiales, tales como documentos fotográficos, manuscritos, planos arquitectónicos, y partituras musicales y bases de datos por mencionar algunos, ya que estos también contribuyen al ciclo de gestión, promoción y visibilidad del conocimiento. En la última década los repositorios tienden a convertirse en plataformas para los datos abiertos. Sin embargo, “los datos, los estándares de evidencia, las formas de representación y las prácticas de investigación están entrelazados. Las diferencias entre comunidades se vuelven aparentes sólo cuando se intenta hacer uso de ellos al combinarlos con datos externos para colaborar entre disciplinas” (Borgman 2015¹).

Si bien se han estudiado los distintos procesos de comunicación científica entre las Humanidades y las Ciencias y el impacto que esto tiene en repositorios, existe menos trabajo en cómo los documentos digitales almacenados en repositorios pueden descomponerse en datos que son utilizados para realizar investigación en las Humanidades y por lo tanto, saber cómo se conforman esos datos y cómo se trabaja con ellos, para que ello sea tomado en consideración en la creación de repositorios (Gómez, Méndez, y Hernández-Pérez 2016).

Estamos viviendo un periodo en donde “los datos” se han convertido en un objeto de estudio de mucha importancia, en particular con la posibilidad del *big data* en donde podemos utilizar una gran cantidad de ellos para realizar análisis y proyecciones de nuevas formas (Borgman 2012). El énfasis en lo “abierto” y la amplia disponibilidad de datos, han generado un interés por la ciencia de datos. Sin embargo, podemos argumentar que los “datos” no son elementos aislados, independientes y objetivos sino que están constituidos a partir de decisiones e interpretaciones humanas y que estas deben contextualizarse para que otros hagan uso de ellos (Lynch 2002). Tanto en las

¹ Todas las traducciones son de los autores.

Ciencias como en las Humanidades, los metadatos acerca de los datos son claves y necesarios pues proveen contexto -en mínima expresión- sin la necesidad de desarrollar una narrativa (Fenner et al. 2019).

Para cada disciplina existen formas particulares de entender qué son datos válidos, su naturaleza y cómo deben ser recabados (Gómez, Méndez, y Hernández-Pérez 2016; Borgman, Wallis, y Mayernik 2012). Las disciplinas humanísticas generalmente no generan datos para investigación, sino que utilizan registros de actividad humana como fuente de estudios, tales como fotografías, cartas, periódicos, libros, artículos, archivos, registros civiles (Borgman 2009). Sin embargo, al digitalizarse estos objetos, tanto a partir de su contenido como de su descripción, pueden ser contruidos conjuntos de datos que permiten hacer una aproximación distinta a estos, transformando el modo como pueden ser analizados tales registros. Para Christof Schöch (2013) “un dato en humanidades puede ser considerado como una abstracción digital, construida selectivamente, manipulable por una máquina, que representa algún aspecto de un objeto dado para el estudio humanístico”.

En este trabajo se plantea que el OCR (*Optical Character Recognition*) de periódicos digitalizados del siglo XIX pueden entenderse como "datos crudos" para la realización de investigación histórica hemerográfica y ofrecemos una reflexión en torno a cómo deben integrarse estos datos al repositorio para poder ser utilizado para estos fines. No pretendemos que se entienda el periódico histórico únicamente como datos, pues como se ha señalado antes los datos, dependiendo de la forma cómo se estructuran, sólo representan un aspecto de ese objeto, nunca su totalidad. Nuestra intención, en cambio, es realizar una reflexión en torno a cómo, al digitalizarse los periódicos y al generarse textos digitales a partir del OCR los investigadores pueden proponer diversas estructuras de datos -por ejemplo, entidades nombradas, palabras más frecuentes, fuente y destino, sólo por citar algunas- para ser procesados a través de una computadora. Esto implica una aproximación epistemológica diferente con respecto a los objetos de estudio, en este caso los periódicos del siglo XIX, pues utiliza representaciones parciales para ser conocidas a través de procesos computacionales. Por eso debemos analizar críticamente esta conversión para entonces contextualizar apropiadamente los resultados que obtenemos cuando realizamos desde cosas sencillas como búsquedas hasta temas más complejos como análisis lingüísticos. El término "datos crudos" puede entenderse como un oxímoron, ya que todos los datos “están cocinados” de alguna forma (Gitelman 2013). Cuando reducimos objetos físicos con cargas culturales e históricos a datos realizamos un proceso interpretativo, y en este trabajo argumentamos que es importante considerar que los “datos” son el resultado de procesos subjetivos y de decisiones humanas que las condiciona.

Materiales y metodología

Las colecciones de periódicos en hemerotecas son una importante fuente para el estudio de la historia, la literatura, la lingüística y otras disciplinas. El manejo de periódicos es notoriamente difícil en el ámbito bibliotecológico tanto en términos de organización documental como en su preservación. En términos de catalogación es

complicado identificar las temáticas de un periódico ya que las noticias y artículos que contiene un número son muchos y variados, por lo que generalmente se identifican solamente el título y la fecha, así como otros datos generales. Los periódicos suelen publicarse diariamente por lo que las hemerotecas generalmente se enfrentan a grandes volúmenes de documentos que son difíciles de almacenar, identificar y dar acceso. Aunado a esto los periódicos están impresos en papel de baja calidad lo que significa que se deterioran rápidamente.

La digitalización de periódicos se empezó a realizar hace ya varias décadas y esto ha permitido un renovado interés por los estudios hemerográficos (Latham & Scholes, 2006). Las colecciones digitales de periódicos permiten a los usuarios consultar los materiales y contribuye a la preservación de los mismos. En general se tiende a estudiar estos repositorios como si fueran sustitutos o suplentes de la versión impresa. La computadora se entiende como una simple ventana al archivo físico y no un sistema de remediación del archivo (Cordell, 2017). Así los investigadores realizan búsquedas y detectan periódicos relevantes pero interpretan los resultados como si hubieran consultado el archivo físico.

Algunos investigadores utilizan el sistema a través de algún tipo de catálogo y encuentran los periódicos por título o por fecha. Sin embargo, la mayoría de los repositorios de periódicos permiten también realizar búsquedas sobre el texto completo, algo que no es posible en el archivo físico, lo cual incrementa el alcance de los materiales que se consultan. Sin embargo, pocos investigadores están conscientes de que las búsquedas sobre el texto completo se hacen sobre archivos de texto plano extraído mediante OCR de las imágenes de los periódicos que ellos están viendo. Las palabras del periódico se convierten en "datos" a partir del momento en que un investigador introduce una palabra o una secuencia de palabras para ser identificadas en el texto del OCR. Sin embargo, salvo algunas excepciones, el investigador basa su marco interpretativo teniendo en mente el archivo físico y no el OCR. Y esa es una cuestión relevante, pues el texto OCR jugará un papel significativo en la obtención de los resultados, en este caso a partir de la búsqueda simple. Por ello es importante empezar a entender el OCR no solamente como un proceso técnico relegado a las máquinas, sino que debemos comprender que se trata de la transformación del objeto que incide directamente en la forma de obtener resultado, por lo que es fundamental saber qué ocurre para contextualizar los resultados desde una perspectiva humanística.

En nuestro caso hemos estudiado la Hemeroteca Nacional Digital de México (HNDM) que contiene más de 7 millones de imágenes y los archivos en XML producto del proceso de OCR correspondiente. El trabajo se realizó en el marco del proyecto de investigación "Oceanic Exchanges: Tracing Global Information Networks in Historical Newspaper Repositories, 1840-1914", un proyecto de minería de datos en repositorios de periódicos digitalizados. El proyecto de investigación OcEx está compuesto por 6 equipos de investigación en 9 países y 11 proveedores de datos (colecciones hemerográficas) y busca identificar y modelar patrones de flujo de información en periódicos del siglo XIX a través de métodos computacionales. Es importante aclarar que OcEx no busca centralizar los repositorios digitales, sino establecer ontologías a partir de la interoperabilidad en las diferentes estructuras de metadatos para evitar las

complicaciones naturales a la multiplicidad de lenguajes participantes en este proyecto, para vincularlos como una red que ofrezca a los investigadores acceso a herramientas analíticas, conectadas a través de bases de datos remotas.

En un principio se consideró que el establecimiento de las ontologías sería relativamente sencillo. Sin embargo, al principio del proyecto de investigación quedó claro que si bien teníamos los datos no estaba suficientemente clara mucha información acerca de ellos: cómo están compuestos, cómo fueron adquiridos, qué tan precisos son, a qué objetos físicos corresponden, qué información metatextual o de metadatos incluyen y qué tan precisa o confiable es. Se realizó una detallada revisión de la documentación disponible, acompañada de una serie de entrevistas a actores claves en la formación de la HNMD así como observaciones del mismo equipo de investigación del proyecto al utilizar los datos de la HNMD para el proyecto.

Lynch (2002) hace una importante diferencia entre “colecciones digitales” y “bibliotecas digitales” que nos interesa retomar aquí. Según Lynch las colecciones digitales son los materiales “en crudo” (*raw materials*) en donde el enfoque está en crear grandes cantidades de contenidos digitales con herramientas muy básica de acceso y descubrimiento. Sobre estos “datos crudos” se realizan algunos trabajos “interpretativos” o de “curaduría” para dar significado y contexto a los contenidos. En cambio, considera que las bibliotecas digitales son “sistemas que hacen que las colecciones digitales cobren vida, los hace accesibles de una forma útil, de tal suerte que sean útiles para nuestro trabajo y las conecta con sus comunidades” (Lynch, trad.propia). En este mismo sentido, para Voutssas (2006) las bibliotecas digitales se caracterizan por contener no solo los materiales digitales sino también ofrecer servicios bibliotecarios (Voutssas 2006). ¿Pero cuáles podrían ser algunos de estos servicios para el caso de la investigación que se interesa por los datos?

Nosotros consideramos la HNMD como un repositorio y no como una biblioteca/hemeroteca digital porque no ofrece servicios bibliotecarios, en especial aquellos dirigidos a la investigación. Solo contiene una colección y, adicionalmente, el grado de catalogación y clasificación es muy básico. Aunado a esto, no brinda acceso directamente a fuentes básicas de datos como los XML formados a partir del OCR. Nuestro equipo de investigación trabajó sólo de manera muy limitada a través de la interfaz de la HNMD para identificar las noticias que podían ser de interés para la investigación. En cambio, utilizó de forma aislada los XML y sus metadatos asociados a estos documentos para realizar la minería de datos que era el objetivo final del proyecto. Es decir, el OCR contenido en los archivos XML fue la fuente de los datos para la investigación tanto en la forma más básica de la búsqueda, como para la extracción, mediante procedimientos computacionales como *sentimental analysis*, entidades nombradas, frecuencias, de la información.

El problema mayor, desde el punto de vista de la formación de colecciones y bibliotecas digitales es el de repensar que todos sus componentes digitales -metadatos, XML, OCR, imágenes, etcétera- tienen valor como representación de aspectos de los objetos digitalizados y, en ese sentido, poseen un alto valor para los usuarios de la colección.

Resultados

La HNDM es un proyecto de digitalización que tiene más de quince años y por lo tanto ha pasado por diversas etapas. La estructura básica sin embargo se ha mantenido. Cada imagen corresponde a una página de un periódico y posteriormente se utilizó un sistema de reconocimiento de caracteres (OCR). El OCR está insertado en un XML que también contiene los metadatos del periódico así como las coordenadas de cada palabra. Esto permite que cuando un usuario realice una búsqueda en la interfaz de la HNDM, la palabra que busca sale resaltada en amarillo en la misma imagen. Si bien es una excelente funcionalidad, contribuye a que el usuario interprete que está haciendo sus búsquedas sobre el periódico digitalizado. El archivo XML en donde se hace la búsqueda está oculto al usuario.

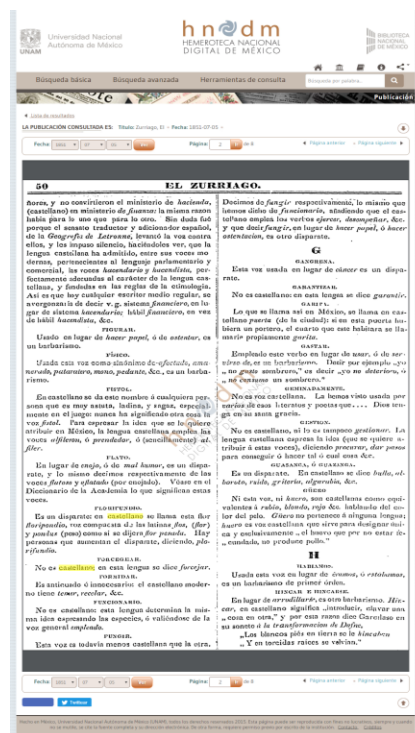


Fig. 1. Búsqueda por palabra en la HNDM. La palabra se resalta en amarillo.

El usuario común no tiene acceso a los archivos XML. Por lo tanto, no tiene la posibilidad de descargar o utilizar los datos sobre los cuales finalmente está haciendo sus búsquedas. Todos los periódicos que se encuentran en el dominio público están disponibles en acceso abierto. La HNDM que está resguardada por la Universidad Nacional Autónoma de México, tiene una política de apoyo a acceso abierto (Gaceta 2015). Los periódicos con acceso restringido solamente se pueden consultar en los equipos de cómputo que se encuentran en el edificio de la Hemeroteca Nacional.

[illegible]

192

También identificamos que la HNDM contiene dos versiones de los archivos XML, los “sucios” que son el resultado del OCR y los “limpios” en donde se realizaron procesos computacionales para eliminar “basura” generada por el OCR, en particular caracteres problemáticos y frecuentes como %, &, *, por mencionar algunos. Estoy incluyó también los acentos. Las búsquedas se realizan sobre los archivos XML “limpios” por lo que las búsquedas ignoran palabras acentuadas o con estos caracteres problemáticos. Si bien, esto no necesariamente es crítico para todas las búsquedas, con algunos términos en donde el acento o algún carácter particular es importante los resultados arrojados por la búsqueda pueden ser problemáticos o deficientes. Esto no está indicado para los usuarios.

También se realizó un análisis de cómo están vinculados los títulos de los periódicos y sus metadatos con las imágenes correspondientes y el XML que lo debe acompañar. En las distintas etapas de la HNDM se migró el sistema en varias ocasiones a nuevas plataformas, tanto en la base de datos como las interfaces para acceder a ellas. Encontramos discrepancias y XMLs faltantes. En ocasiones entonces un usuario puede encontrar un periódico y su imagen y asume incorrectamente que el sistema de búsqueda de texto completo las incorpora. Sin embargo, como la búsqueda se realiza sobre el OCR, si el XML correspondiente no existe en el sistema, entonces no los incluye. Esto es algo que pasa desapercibido por el usuario.

Conclusiones

El OCR de periódicos digitalizados del siglo XIX presenta varios problemas de fidelidad con el original sin embargo, esto no significa que no puede ser utilizado para realizar estudios de minería de datos. Si pensamos en el acceso abierto, es importante que no sólo tengamos acceso a las imágenes de los periódicos pero si queremos poder procesarlos y entenderlos, también se requiere el acceso a los datos crudos, en este caso los archivos XML con el OCR y los metadatos.

Uno de los principales problemas para el uso de estos repositorios con fines de investigación es la falta de descripción adecuada de los datos, su contexto y la relación con los metadatos y la interfaz de búsqueda. Para lograr esto es necesario aproximarse a la digitalización, o mejor dicho, a los datos resultantes de una digitalización, desde una perspectiva crítica e interpretativa y no únicamente tecnológica. El OCR derivado de un texto histórico constituye una nueva edición del texto (Cordell, 2017, p.196). Por lo tanto consideramos que los repositorios que almacenan periódicos digitalizados deben de incluir una descripción hemerográfica no solo de la colección impresa sino también de la colección digitalizada, en donde se realice un trabajo de descripción bibliográfica para “dar cuenta de las fuentes, tecnologías y realidades sociales de su creación de manera que sus posibilidades y limitaciones sean más fácilmente visibles y estén disponibles para la crítica” (Cordell, 2017).

El uso de repositorios para custodiar colecciones de datos crudos permite compartir y reutilizar esta información. En este caso de estudio buscamos entender el OCR de periódicos digitalizados como datos crudos para la investigación histórica. Esto constituye un primer acercamiento y reflexión a cómo este tipo de datos deben de ser

almacenados en repositorios para que sean adecuadamente interpretados y utilizados en la investigación.

Referencias

- Gaceta (2015). Lineamientos generales para la política de acceso abierto de la UNAM, 10 de septiembre 2015.
- Latham, S., & Scholes, R. (2006). The Rise of Periodical Studies. *PMLA*, 121(2), 517-531. <https://doi.org/10.1632/003081206X129693>
- Borgman, Christine L. 2009. «The Digital Future is Now: A Call to Action for the Humanities» 3 (4). <http://www.digitalhumanities.org/dhq/vol/3/4/000077/000077.html>.
- . 2012. «The Conundrum of Sharing Research Data». *Journal of the American Society for Information Science and Technology* 63 (6): 1059-78. <https://doi.org/10.1002/asi.22634>.
- Borgman, Christine L. 2015. *Big Data, Little Data, No Data Scholarship in the Networked World*. Cambridge, Massachusetts: The MIT Press.
- Borgman, Christine L., Jilian Wallis, y Matthew Mayernik. 2012. «Who's Got the Data? Interdependencies in Science and Technology Collaborations». SSRN Scholarly Paper ID 2089165. Rochester, NY: Social Science Research Network. <https://papers.ssrn.com/abstract=2089165>.
- Cordell, Ryan. 2017. «“Q i-Jtb the Raven”: Taking Dirty OCR Seriously». *Book History* 20 (1): 188-225. <https://doi.org/10.1353/bh.2017.0006>.
- Fenner, Martin, Mercè Crosas, Jeffrey S. Grethe, David Kennedy, Henning Hermjakob, Phillippe Rocca-Serra, Gustavo Durand, et al. 2019. «A data citation roadmap for scholarly data repositories». *Scientific Data* 6 (1). <https://doi.org/10.1038/s41597-019-0031-8>.
- Gitelman, Lisa, ed. 2013. «*Raw data*» is an oxymoron. Infrastructures series. Cambridge, Massachusetts ; London, England: The MIT Press.
- Gómez, Nancy-Diana, Eva Méndez, y Tony Hernández-Pérez. 2016. «Social Sciences and Humanities Research Data and Metadata: A Perspective from Thematic Data Repositories». *El Profesional de La Información* 25. <http://eprints.rclis.org/30054/>.
- Lynch, Clifford. 2002. «Digital Collections, Digital Libraries and the Digitization of Cultural Heritage Information». *First Monday* 7 (5). <https://doi.org/10.5210/fm.v7i5.949>.
- Schöch, Christof. 2013. «Big? Smart? Clean? Messy? Data in the Humanities». *Journal of Digital Humanities* 2 (Summer). <http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/>.
- Voutssás, Juan. (2006) Bibliotecas y publicaciones digitales, Universidad Nacional Autónoma de México. Disponible en: http://ru.iibi.unam.mx/jspui/bitstream/IIBI_UNAM/L67/1/bibliotecas_y_publicaciones_digtales.pdf

A implantação da Base de Dados Científicos (BDC) da Universidade Federal do Paraná (UFPR)

Karolayne Costa Rodrigues de Lima¹, Paula Carina de Araújo², Janete Saldanha Bach Estevão³

¹ Bibliotecária na Universidade Federal do Paraná. Mestranda em Gestão da Informação pela Universidade Federal do Paraná (UFPR). Email: karolbraun@gmail.com

² Bibliotecária na Universidade Federal do Paraná. Doutora em Ciência da Informação pela Universidade Estadual Paulista Júlio de Mesquita (UNESP). Email: paulacarina@ufpr.br

³ Bibliotecária na Universidade Federal do Paraná. Doutoranda em Tecnologia pela Universidade Tecnológica Federal do Paraná (UTFPR). Email: janeteest@gmail.com

Resumo

A Base de Dados Científicos (BDC) da Universidade Federal do Paraná (UFPR) visa disponibilizar os dados científicos utilizados das pesquisas que são publicadas pela comunidade da UFPR em teses, dissertações, artigos de revistas e outros materiais bibliográficos. O objetivo desta comunicação é descrever o processo de implantação da BDC no âmbito da UFPR, desde as escolhas estratégicas até o seu efetivo uso. A implantação da BDC passou por algumas etapas como: a seleção do software, a seleção de requisitos, a seleção do padrão de metadados, o estabelecimento da metodologia para formular as diretrizes para o depósito dos dados, a especificação dos critérios para a criação do plano de gestão de dados, o registro no serviço indexador de repositórios de dados, o Re3data.org e a divulgação para a comunidade acadêmica. No decorrer do processo de planejamento e implantação da BDC, destacam-se alguns desafios: os custos com atribuição de DOI para cada *dataset*, o arranjo de características comuns para base de dados científicos que atendam aos contextos disciplinares distintos em uma única plataforma, a capacitação contínua dos usuários, a estrutura para oferta de serviços e a capacidade de suporte, além do entendimento do processo de produção e da evolução científica no contexto da ciência aberta.

Palavras-chave: Dados de pesquisa. Dados abertos. Repositório de dados. Base de Dados Científicos da Universidade Federal do Paraná (BDC/UFPR).

Abstract

The Scientific Database (CDB) of the Federal University of Paraná (UFPR) aims to make available the scientific data used from the researches that are published by the UFPR community in theses, dissertations, journal articles and other bibliographic materials. The thematic axis of this proposal is inserted in repositories of research data, specifically in the management and curation of data repositories. The purpose of this communication is to describe the process of implementation of the BDC within the scope of UFPR, from the strategic choices to its effective use. The implementation of BDC has undergone several steps: software selection, selection of requirements, selection of the metadata standard,

establishment of the methodology for formulating the guidelines for data storage, specification of the criteria for the creation of the plan data management, registration in the data repository index service, Re3data.org and dissemination to the academic community. During the process of planning and implementation of the BDC, some challenges are highlighted: the costs with DOI assignment for each dataset, the arrangement of common characteristics for scientific data bases that attend to the different disciplinary contexts in a single platform, the structure for offering services and support capacity, as well as the understanding of the production process and the scientific evolution in the context of open science.

Keywords: Scientific data. Open data. Data repository. Research data database of Paraná Federal University (BDC/UFPR).

Introdução

A partir dos anos 2000, as políticas que advogam pelo acesso aberto da comunicação da pesquisa científica, financiada com recursos públicos, têm impulsionado a difusão e o compartilhamento dos dados produzidos nos processos de investigação (CREASER, 2011, p. 56). Diversas instituições de fomento têm estabelecido políticas mais específicas de acesso aos dados de pesquisa daqueles que recebem recursos público (BERLIN, 2003; NATIONAL SCIENCE BOARD, 2005; NATIONAL SCIENCE FOUNDATION, 2007, 2010; OECD, 2007). Outra crescente exigência de depósito de dados tem sido protagonizada pelas revistas científicas, o que torna um pré-requisito aos pesquisadores estarem preparados para atuarem no cumprimento dessas regras. Um estudo sobre a atuação das revistas científicas no contexto de dados abertos mostrou que, dentre as 77 revistas brasileiras da área das ciências disponíveis no *Directory of Open Access Journals* (DOAJ), 15 delas já exigem o depósito de dados, sendo uma delas um periódico exclusivo de dados. Na área de Medicina, das 139 revistas indexadas, 71 fazem a mesma exigência (CARVALHO, 2016).

Buscando o contínuo alinhamento e a integração das melhores práticas em Ciência Aberta, a Universidade Federal do Paraná (UFPR), por meio de uma parceria entre o Centro de Computação Científica e Software Livre (C3SL) e o Sistema de Bibliotecas (SiBi) da UFPR, tendo a frente uma equipe multidisciplinar, planejou, desenhou e implementou, entre setembro de 2017 a janeiro de 2018, o primeiro repositório de uma universidade pública brasileira para dados científicos, a Base de Dados Científicos da Universidade Federal do Paraná (BDC/UFPR).

Portanto, o objetivo desta comunicação é descrever o processo de planejamento, decisões estratégicas e implantação da BDC/UFPR. Este projeto foi motivado pela contínua busca da excelência desta universidade em acompanhar a tendência mundial de planejamento, gestão, produção, organização, armazenamento, disseminação e reuso de dados de pesquisa. A disponibilização dos dados de pesquisa contribui para a transparência e otimização da produção científica por meio do reuso dos conjuntos de dados e a possibilidade de novas análises e abordagens.

Dados científicos de pesquisa

A mudança do paradigma científico provocada pela confluência das tecnologias de informação e comunicação no fazer científico tomou força a partir da metade para o final do

século XX e foi caracterizada por uma série de movimentos de abertura na prática e na comunicação da ciência. Como exemplo dessa abertura, aponta-se para o movimento do *software* livre de código aberto que permitiu o desenvolvimento de sistemas abertos (repositórios de dados); movimento de acesso aberto à informação científica na disponibilização por meio dos primeiros periódicos de acesso aberto; o fomento no desenvolvimento de recursos e práticas educacionais abertos, metodologia aberta, dados abertos, entre outros.

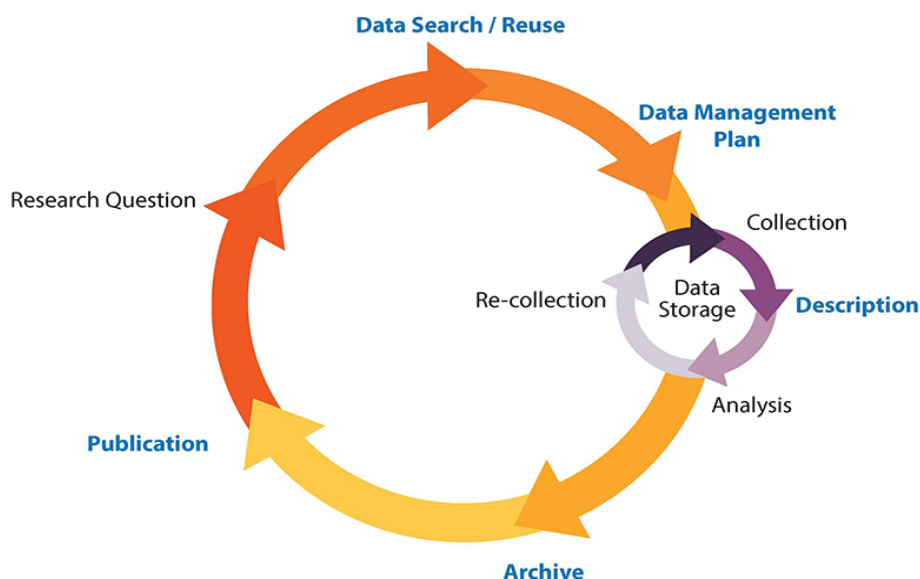
Nesse contexto de abertura, destacam-se ainda os dados científicos de pesquisa como uma unidade nuclear primordial do processo de investigação. Os dados científicos de pesquisa “referem-se ao material de fato registrado, comumente aceito na comunidade científica como necessário para validar os resultados da pesquisa” (THE ENGINEERING AND PHYSICAL SCIENCES RESEARCH COUNCIL, 2018, tradução nossa).

A considerar que o dado dentro de um projeto de pesquisa caracteriza-se como insumo ou produto, os conjuntos de dados (*datasets*) científicos têm características e particularidades dependendo de cada área do conhecimento e cobrem uma ampla gama de tipos de registros, podendo ser estruturados e armazenados em vários formatos de arquivos.

A pensar o contexto dos dados científicos de pesquisa dentro da perspectiva da comunicação científica, destacam-se certos princípios atrelados à abertura dos dados, os quais são: a **publicidade**, conferindo maior visibilidade aos pesquisadores; a indução da **colaboração** em rede e transparência dos dados utilizados para os resultados; a possibilidade de **reuso** de dados em novas conexões; a **aceleração** da produção de novas pesquisas; o atendimento às **regras de financiadoras** de pesquisa; a promoção da **reprodutividade**, a verificabilidade para garantir boa prática científica, evitando fraudes; e o **acesso** à pesquisa de importância social e maior consciência dos desafios da sociedade.

A pesquisa em torno dos dados científicos é ampla e não se restringe apenas ao seu armazenamento em repositório de dados que é o foco deste trabalho. Todavia, é relevante ilustrar o ciclo de gestão de um dado científico de pesquisa, conforme figura 1 abaixo:

Figura 1: Ciclo de gestão de dados científicos



Fonte: Digital Curation Center (2018).

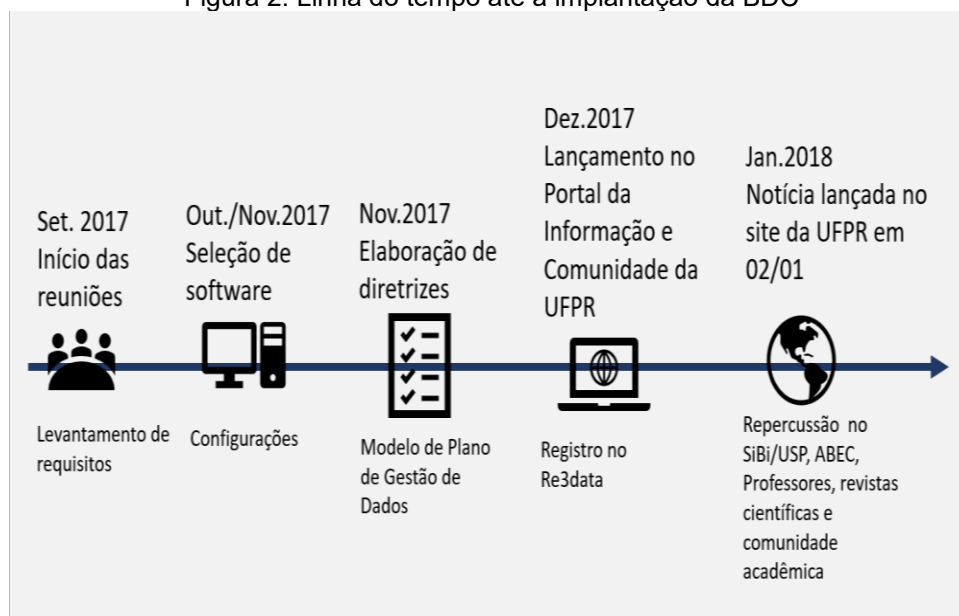
Base de Dados Científicos da Universidade Federal do Paraná (UFPR)

A Base de Dados Científicos (BDC) da Universidade Federal do Paraná (UFPR) foi lançada em dezembro de 2017 e é fruto do trabalho de uma equipe multidisciplinar do Centro de Computação Científica e Software Livre (C3SL) e o Sistema de Bibliotecas (SiBi) da UFPR. A BDC visa reunir os dados científicos utilizados nas pesquisas que foram publicadas pela comunidade da UFPR em teses, dissertações, artigos de revistas, e outros materiais bibliográficos. Trata-se de um repositório de caráter multidisciplinar para o depósito de dados científicos.

Além disso, a BDC é um serviço inovador que acompanha a tendência mundial de planejamento, gestão, produção, organização, armazenamento, disseminação e reuso de dados de pesquisa. A disponibilização dos dados de pesquisa contribui para a transparência e otimização da produção científica por meio do reuso dos conjuntos de dados e a possibilidade de novas análises e abordagens.

A implantação da BDC/UFPR ocorreu em diferentes etapas, tais como: a seleção do software; a seleção de requisitos; a seleção do padrão de metadados, o estabelecimento das metodologias para formular as diretrizes para o depósito dos dados, a especificação a seleção dos requisitos e dos critérios para a criação do plano de gestão de dados, o registro no serviço indexador de repositórios de dados, o Re3data.org e a divulgação para a comunidade acadêmica. Cada uma das etapas será descrita a seguir.

Figura 2: Linha do tempo até a implantação da BDC



Fonte: Setenareski; Estevão; Lima (2019)

- **Seleção do Software:** foram testados os serviços de hospedagem de dados e os *softwares* Zenodo, Fedora, Invenio, Open Science Framework e Ckan no estudo de viabilidade da ferramenta a ser utilizada. Porém, optou-se por manter o Dspace, *software* livre que com o qual a equipe está familiarizada, uma vez que já é usado para o Repositório Digital Institucional (RDI) da UFPR, e que atendeu aos pré-requisitos estabelecidos pela equipe, além das funcionalidades inerentes ao depósito de dados.



- **Seleção de Requisitos:** Foi utilizada a matriz de casos de uso e de requisitos funcionais, da Research Data Alliance (RDA)¹, selecionados para plataformas de repositório de dados de pesquisa (<https://goo.gl/owqXHH>).
- **Seleção de Padrão de Metadados:** o Diretório de Padrões de Metadados da RDA (<https://goo.gl/DBjbvV>) foi utilizado para um maior entendimento do que estava sendo adotado por outras instituições. Porém, dentre os repositórios registrados no Registry of Research Data Repositories (re3data.org²), o mais utilizado era o padrão Dublin Core, sendo este também adotado para a BDC/UFPR.
- **Metodologia para estabelecer as Diretrizes para depósito de dados:** a partir do método da análise de conteúdo, foi realizado um estudo das políticas para depósito de dados de instituições internacionais, que já estavam em estágio de maior maturidade em Repositórios de Dados Científicos: Revista Nature, Repositório Dryad, Repositório ICPSR, Repositório PLOS One, Repositório da Universidade de Edimburgo e Repositório Universidade do Texas. Também avaliou-se a Política do Inter-university Consortium for Political and Social Research (ICPSR³) e manteve-se o alinhamento com as orientações das recomendações do guia geral do Grupo de trabalho para políticas de repositório da RDA, o *Practical Policy Working Group*. As [diretrizes](#) apresentam os principais conceitos, orientações para elaboração do plano de gestão de dados; a estrutura e funcionamento da BDC/UFPR, incluindo orientações do processo de submissão e de direitos autorais.
- **Seleção de requisitos e formulação do Plano de Gestão de Dados:** o Plano de Gestão de Dados da BDC/UFPR foi resultado de uma avaliação de ferramentas automatizadas para geração de Plano de Gestão de Dados, tais como: - o DPMTTool (<https://dmptool.org>); - o DCC (<https://dmponline.dcc.ac.uk>). Também foram avaliados modelos de Planos de Gestão de dados e outros padrões relevantes como: - o DMP Horizon 2020 (<https://goo.gl/nfsEKf>); - o DMP ICPSR (<https://goo.gl/TyrNLZ>). Essas ferramentas e modelos fundamentam as recomendações dos elementos principais que deveriam compor o Plano de Gestão de Dados da UFPR. O Plano de Gestão de Dados da BDC se propõe a descrever como os dados científicos serão tratados durante a pesquisa e, também, após a sua conclusão. É formado pelas seguintes seções: - identificação do pesquisador/grupo de pesquisa; - identificação da pesquisa; - descrição dos dados; - outras informações / metadados; - anuência das Diretrizes da BDC
- **Registro no Re3data.org:** prodeceu-se também o registro da BDC/UFPR no Registry of Research Data Repositories, principal indexador dos repositórios de dados em vários países.
- **Divulgação e Repercussão:** o lançamento da BDC/UFPR ocorreu no Portal da

¹ A RDA foi criada em 2013 com financiamento da Comissão Europeia, do NSF (Estados Unidos) e do departamento de inovação do governo australiano, com o objetivo de construir a infraestrutura social e técnica para permitir o compartilhamento aberto de dados científicos nas mais diversas áreas do conhecimento. Conta com mais de 5900 membros de 129 países (dados de agosto/17) (<https://www.rd-alliance.org/node/51727>). Apenas 2% dos membros da organização estão localizados na América Latina.

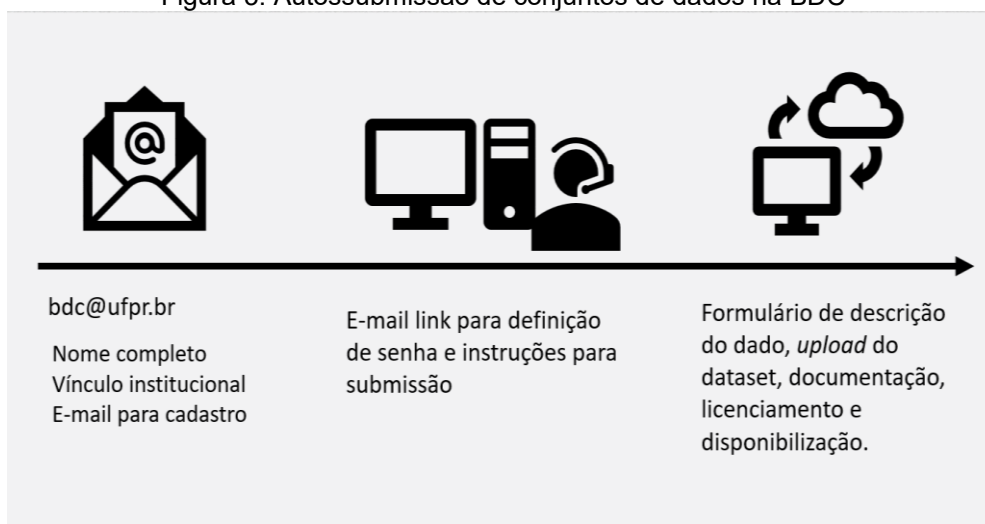
² A maior parte desses recursos indexados no re3data.org são repositórios para depósito de dados de instituições de pesquisas e de universidades. A pesquisa no re3data.org pode ser feita por um conjunto de critérios, como tipo de acesso, tipo de conteúdo, estrutura de dados, padrão de metadados utilizado, licença de uso de dados, país de origem, idioma, tipo de versionamento, área do conhecimento, *software* utilizados para o repositório, dentre outros.

³ O ICPSR é consórcio internacional com mais de 750 instituições afiliadas no mundo que está na liderança em gestão de dados em Ciências Sociais, disponibilizando mais de 250 mil arquivos de dados científicos.

UFPR e houve a divulgação nas redes sociais. A Repercussão foi imediata, sendo compartilhado por Sistemas de Bibliotecas de outras universidades do país, despertando interesse também da Associação Brasileira de Editores Científicos (ABEC), de professores da universidade e externos a ela e de revistas científicas externas à UFPR. Houve grande interesse da comunidade acadêmica e, a equipe de suporte e orientação aos usuários interessados quando ao processo de depósito e documentação de *datasets* realizou 135 atendimentos apenas no primeiro semestre de 2018.

Após a implantação da BDC, há um grande esforço para a divulgação do serviço e esclarecimentos aos dos pesquisadores da UFPR quanto à importância e as vantagens de compartilharem seus dados de pesquisa desenvolvidas no âmbito da UFPR. A disponibilização dos dados se dá por meio da autossubmissão na comunidade específica dentro do repositório. A Figura 2 esclarece o processo de autossubmissão.

Figura 3: Autossubmissão de conjuntos de dados na BDC



Fonte: Setenareski; Estevão; Lima (2019)

A maioria dos atendimentos aos pesquisadores se deu por e-mail. Inicialmente, o pesquisador envia um e-mail à equipe da BDC solicitando a divulgação dos seus dados/arquivos na BDC. Em seguida, é enviado um e-mail ao pesquisador com um texto informativo sobre objetivo da BDC para que o mesmo informe se os dados/arquivos que tem interesse em publicar na BDC são caracterizados como dados produzidos nos processos de investigação. Em caso positivo, o pesquisador precisará enviar um e-mail para a equipe da BDC com: nome completo, e-mail de contato, matrícula UFPR, vínculo com a UFPR (aluno, professor, outro), e informar a qual trabalho (artigo, tese, dissertação, monografia especialização/graduação) os seus dados/arquivos são referentes. Caso o pesquisador não tenha mais vínculo com a UFPR, informará à equipe da BDC: nome completo, e-mail de contato, matrícula UFPR, programa de (pós-) graduação da UFPR em que se formou, a qual trabalho (artigo, tese, dissertação, monografia especialização/graduação) os seus dados/arquivos são referentes, e a data de publicação/defesa deste trabalho.

Após a análise destas informações a equipe da BDC cadastra o pesquisador na BDC para que o mesmo proceda com a submissão dos seus dados/arquivos. O processo ocorre

por autosubmissão, sendo que o pesquisador recebe um tutorial que o auxilia no preenchimento do formulário de descrição dos dados, *upload* do *dataset*, documentação, licenciamento e disponibilização. Durante o processo de contato entre o pesquisador e a equipe da BDC é informado a possibilidade de criação de coleção específica para grupo de pesquisa da UFPR. Desta forma a BDC possibilita que mais de um pesquisador seja vinculado a esta coleção e possa submeter os dados/arquivos do grupo de pesquisa em um único lugar e em conjunto com os outros pesquisadores atuantes neste grupo de pesquisa.

Atualmente, há uma limitação de 2 gigabytes para o tamanho do *dataset* a ser depositado, dessa forma, *datasets* maiores são submetidos pela equipe da BDC. Também percebe-se que uma parte expressiva dos pesquisadores ainda não está habituada a elaborar um Plano de Gestão de Dados antes de iniciar a sua pesquisa, o que também representa um fator limitador. Compreende-se que é necessária uma mudança no processo de pesquisa no que diz respeito à disponibilidade dos dados, pois, ainda não há em todas as áreas do conhecimento, a ação do compartilhamento e do reuso dos dados de pesquisa. Em outros países, essa prática foi iniciada por força de leis e políticas institucionais que requerem o depósito dos dados ao final da pesquisa financiada com recursos públicos.

Durante os atendimentos, percebeu-se que muitos pesquisadores têm dificuldade em diferenciar a divulgação dos seus dados produzidos nos processos de investigação, da divulgação do trabalho ao qual estes dados estão vinculados. Este cenário, mostra a necessidade de capacitar os pesquisadores quanto ao entendimento da plataforma BDC. Para dirimir esta dificuldade, a equipe da BDC produziu uma Frequently Asked Questions (FAQ), em que detalha informações essenciais para o entendimento da BDC. Além disso, a equipe da BDC analisa os dados/arquivos submetidos na BDC e verificam se os mesmos estão adequados ao objetivo da plataforma. Em caso negativo entram em contato com o pesquisador para rever os dados submetidos.

Considerações Finais

Um projeto dessa natureza demanda o trabalho engajado em equipe, de pessoas de formações multidisciplinares. Bibliotecários, analistas e técnicos em Tecnologia da Informação e da Computação e professores contribuíram para a viabilidade deste serviço. Em termos de desafios, destacam-se os custos com atribuição de DOI para cada *dataset* e o arranjo de características comuns para base de dados científicos que atendam aos contextos disciplinares distintos em uma única plataforma. A capacitação contínua dos usuários, visto que é uma atividade recente no Brasil, principalmente em relação ao letramento no gerenciamento dos dados de pesquisa, também é um ponto de atenção. A estrutura para oferta de serviços e a capacidade de suporte, mantendo uma equipe multidisciplinar no atendimento constantemente atualizada, é um dos aspectos de maior impacto ao valor percebido dos serviços pela comunidade. Por fim, o entendimento do processo de produção científica e da evolução da Ciência no contexto do *e-Science*, são determinantes para o sucesso dos Repositórios de dados como um serviço aderente e relevante às necessidades dos pesquisadores do Século XXI.

Referências

AMORIM, Ricardo Carvalho; CASTRO, João Aguiar; SILVA, João Rocha da; RIBEIRO, Cristina. A Comparison of Research Data Management Platforms: Architecture, Flexible

Metadata and Interoperability. **Universal Access in the Information Society**, v. 16, n. 4, nov. 2017, p. 851–62. <https://doi.org/10.1007/s10209-016-0475-y>. Acesso em: 19 mai. 2019.

ARAUJO, Alessandra Belezia; FERREIRA, Elisabete; FÜHR, Fabiane; MOREIRA, Fernando Cavalcanti; ESTEVÃO; Janete Saldanha Bach; LIMA, Karolayne Costa Rodrigues de; ARAÚJO, Paula Carina de; SETENARESKEI, Ligia Eliana; GONÇALVES, Lucas Henrique; SCHMITZ, Rafaela Paula. **Diretrizes da Base de Dados Científicos da Universidade Federal do Paraná**. Curitiba, PR: SiBi/UFPR, 2018. Disponível em: https://www.portal.ufpr.br/documentos/BDC/diretrizes_BDC.pdf. Acesso em: 10 abr. 2019.

BERLIN Declaration on open access to knowledge in the sciences and humanities. Berlin, 2003. Disponível em: http://www.zim.mpg.de/openaccess-berlin/berlin_declaration.pdf. Acesso em: 08 mai. 2019.

CARVALHO, Teila Oliveira. A Influência das Revistas Científicas de Acesso Aberto para o Depósito e Publicação dos Dados de Pesquisa. In: 7. Conferência Luso-Brasileira sobre Acesso Aberto. **Anais....** 2016. Disponível em: <https://goo.gl/kjZhCM>. Acesso em: 11 mai. 2019.

CREASER, Claire. Scholarly communication and access to research output. In: EVANS, WENDY, BAKER, DAVIS. **Libraries and Society**: role, responsibility and future in an age of change. EBSCO Publishing. 2011. p. 53-66.

DIGITAL CURATION CENTER. **Research data life cycle**. Disponível em: <http://www.dcc.ac.uk/>. Acesso em: 01 mai. 2019.

NATIONAL SCIENCE BOARD. **Long-lived digital data collections: enabling research and education in the 21st century**. National Science Foundation, Sept. 2005. Disponível em: <http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf>. Acesso em: 01 mai. 2019.

NATIONAL SCIENCE FOUNDATION. **Cyberinfrastructure Vision for 21st Century Discovery**. 2007. Disponível em: <https://goo.gl/AEFEJ4>. Acesso em: 10 abr. 2019.

NATIONAL SCIENCE FOUNDATION. **Dissemination and sharing of research results**. 2010. Disponível em: <https://goo.gl/E6nsXV>. Acesso em: 10 mai. 2019.

ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **OECD principles and guidelines for access to research data from public funding**. Paris: Organization for Economic Co-operation and Development, 2007. Disponível em: <http://www.oecd.org/sti/sci-tech/38500813.pdf>. Acesso em: 31 abr. 2019.

SETENARESKEI, Ligia E.; Estevão, Janete Bach; Lima, Karolayne C. R. **Ciência aberta: os movimentos que a impulsionam, como se relacionam e como se inserem no mercado das publicações científicas**, 19 mai. 2019. 29 slides.

Re-escritura da comunicação científica: um caminho para a ciência aberta

Cristina Marques Gomes¹, Adriana Carla S. Oliveira²

¹ Departamento de Ciências da Comunicação - Universidade Federal de Santa Maria (UFSM). Email: cristina@alumni.usp.br

² Professora Colaboradora da Universidade Federal do Rio Grande do Norte e da Escola de Magistratura do Rio Grande do Norte. Email: adrianacarla.a@gmail.com

Resumo

Tradicionalmente, o tripé “pesquisa, sistema e sociedade” (PSS) tem sido a base da concepção de comunicação científica (CC). Ao longo das últimas décadas, porém, diversas transformações de ordem tecnológica vêm ocorrendo no ambiente da CC, assim como na sociedade em geral. Nesse sentido, a presente pesquisa, emerge a partir dos seguintes problemas de investigação: a noção de CC relacionada ao tripé PSS ainda é condizente e/ou suficiente para explicar o fenômeno na atualidade? Caso contrário, uma noção contemporânea da CC deveria incorporar quais outros estratos? E como todos esses aspectos estão diretamente relacionados à concepção de “ciência aberta”? Como forma de responder a tais questões, este estudo apresenta os resultados parciais de uma pesquisa finalizada em 2012 - por meio de múltiplos procedimentos metodológicos qualitativos e quantitativos - com doze projetos do Sétimo Programa-Quadro (FP7) da União Europeia (UE) e indica os caminhos que estão sendo trilhados na pesquisa atual das autoras sobre os projetos do Horizon 2020 também da UE. À luz das discussões sobre a CC aponta alguns dos princípios básicos que poderão nortear uma possível *re-escritura* da comunicação científica em consonância com uma ciência aberta.

Palavras-chave: comunicação científica, ciência aberta, União Europeia.

Abstract

Traditionally, the tripod “research, system and society”(RSS) has been a basis of the Scientific Research Network. During the last decades, however, various transformations of technology order to obtain the environment of the Scientific Communication (SC), as well as society in general. In this sense, the present research emerges as a research problems: does a notion of SC related to the RSS tripod still be adequate and/or sufficient to explain the phenomenon at present? If not, should a contemporary notion of Scientific Communication be incorporated into other strata? And how all aspects are related to the theme “open science”? As an answer to these questions, this study presents the partial results of a research completed in 2012 - through multiple qualitative and quantitative

methodological procedures - with twelve projects of the Seventh Framework Program of the European Union (EU) and indicates the paths being pursued in the authors current research on Horizon 2020 projects also from the European Union. The light of the discussions in Scientific Communication presents some of the basic principles that allow a possible rewriting of SC on the basis of an open science.

Keywords: scientific communication, open science, European Union.

Introdução

A presente pesquisa é fruto de uma problemática que é envolta por diversos arranjos de distintas naturezas. Observando a literatura existente sobre a comunicação científica (CC), por exemplo, nos deparamos com pesquisas que vão desde perspectivas específicas, como a revisão por pares (*peer review*) no âmbito dos periódicos científicos, passando por questões associadas aos editores e por proposições que entrelaçam um conjunto de atores sociais diferentes como, no caso, do *open access* (OA) e outros. Diante, pois, de toda essa multiplicidade - de linhas de investigação e intervenientes - estaremos, aqui, considerado, e refletindo sobre, a comunicação científica (e seus fluxos/sistemas/processos), como uma área complexa e abrangente que, tradicionalmente, está relacionada¹: à “pesquisa” - quando da elaboração de uma investigação, via a comunicação entre os pares em todos os níveis; ao “sistema”, ou seja, a informação que flui de e entre os “intermediários da cadeia de comunicação” (MEADOWS, 1974) - editoras, bibliotecas, agências de financiamento, dentre outros; e à “sociedade” em si - quando do compartilhamento do conhecimento científico na ótica da comunicação da ciência.

A comunicação científica, em vista disso, é um tema multifacetado² que incorpora um conjunto de processos e variáveis associados às inúmeras teorias, pesquisas e projetos em âmbito global. Para Hurd (1996), investigadora, por exemplo, a CC apresenta várias interações com diferentes organizações a partir do momento que envolve a produção, organização e disseminação do conhecimento; já a Microsoft Research, que se dedica a pesquisa básica e aplicada em Ciências da Computação e Engenharia de *Software*, considera o “ciclo de vida da CC” composto por 4 fases³ (*Data Collection, Research & Analysis; Authoring; Publication & Dissemination; e Storage, Archiving & Preservation*); e o LiquidPub⁴, um projeto financiado pelo Sétimo Programa-Quadro (FP7)

¹ Ou, do mesmo modo, segundo Bjork (2007): a “Perform the Research” para a “Pesquisa”; “Fund R&D” e “Communicate the Results” para “Sistema” e “Apply the Knowledge” para “Sociedade” ou, ainda, de acordo com Garvey e Griffith (1979) a “produção, disseminação e uso da informação científica”.

² Fonte: ODLYZKO, Andrew. *The future of scientific communication*. Disponível em: <<http://www.dtc.umn.edu/~odlyzko/doc/future.scientific.comm.pdf>>, acessado em 18 de outubro de 2010.

³ Fonte: Microsoft Research. Disponível em: <http://www.microsoft.com/mscorp/tc/scholarly_communication.msp>, acessado em 13 de julho de 2010.

⁴ Disponível em: <<http://liquidpub.org/>>.

da União Europeia, encara a CC pela ótica de como “o conhecimento científico é produzido, divulgado, avaliado e consumido”. Essa pluralidade de olhares, a partir de diferentes atores, para com o mesmo objeto, coaduna-se com a contemporaneidade e junta-se com a percepção de que o mundo atual está em transição e que é complicado “ignorar” tal fato. As novas tecnologias (NTs) “estão aí” também, disponíveis, atropelando, embaralhando, modificando as estruturas, as dinâmicas da ciência, as formas “de pensar” e, conseqüentemente, a CC e, em paralelo, qualquer conjuntura ou perspectiva de uma ciência aberta.

Estamos imersos, pois, numa sociedade pós-moderna, rodeados por um ambiente tecnológico e cuja ciência, além da especialização, passa por um processo de *desdogmatização*⁵, ou seja, temos diante de nós um quadro complexo, cheio de referências e estudos anteriores e difícil de ser mensurado e, que, nos últimos anos, foi impregnado por uma série de transformações de diversas ordens.

Temos, portanto, as seguintes indagações: mediante às transformações de ordem tecnológica ocorridas no ambiente da CC e na sociedade em geral ao longo das últimas décadas, a noção de comunicação científica considerada a partir de sua acepção tradicionalmente relacionada ao tripé “pesquisa, sistema e sociedade” ainda é condizente e/ou suficiente para explicar o fenômeno na atualidade? Caso contrário, uma noção contemporânea da CC deveria incorporar quais outros estratos? E como todos esses aspectos estão diretamente relacionados à concepção de “ciência aberta”? Como forma de responder a tais questões, este estudo apresenta os resultados parciais de uma pesquisa finalizada em 2012 - por meio de múltiplos procedimentos metodológicos qualitativos e quantitativos - com doze projetos do Sétimo Programa-Quadro (FP7) da União Europeia (UE) e indica os caminhos que estão sendo trilhados na pesquisa atual das autoras sobre os projetos do Horizon 2020 também da UE. À luz das discussões sobre a CC aponta alguns dos princípios básicos que poderão nortear uma possível *re-escritura* da comunicação científica em consonância com uma ciência aberta.

Hipóteses e outras colocações metodológicas derivadas

Iniciamos nossa trajetória com um questionamento primeiro que é: a contemporaneidade convive com a “tecnologia” e esta é “somente” uma força que atua sobre o(s) sistema(s), como qualquer uma das outras, ou a “tecnologia” (re)criou um outro ambiente para a CC? Acreditamos, por certo, numa terceira via, entrelaçando as duas opções, ou seja, um macroambiente tecnológico que envolve a CC contemporânea e a “tecnologia em si” como uma força que atua sobre o(s) sistema(s), processos e todos os outros elementos da CC e que está continuamente se “confrontando” com os limites do que é ou não científico e, nesse intento, será que o macroambiente tecnológico ou a tecnologia, enquanto força, ou as ferramentas e aplicativos derivados da mesma, na contemporaneidade, ou ainda, a tecnologia em associação com as outras forças do macroambiente, consegue “minimizar” os aspectos negativos atrelados ao “poder”

⁵ Expressão utilizada por Sousa (1993).

estendendo, por consequência, a CC, enquanto campo de atuação (processos) e propósitos, sem “subtrair” seu “ pilar” principal, ou seja, sua “cientificidade”?

Esta questão, que não se encerra aqui, é, de forma sintética, assim traduzida no contexto das hipóteses básicas da pesquisa: a noção de comunicação científica considerada a partir de sua acepção tradicionalmente relacionada ao tripé “pesquisa, sistema e sociedade” não é condizente, tampouco suficiente, na contemporaneidade, para explicar o fenômeno. Nesse sentido, a hipótese subjacente é que uma noção contemporânea da CC deve incorporar o estrato tecnológico e incluir uma “relação positiva” entre as variáveis “tecnologia” e “cientificidade” e uma “relação negativa” entre as variáveis “tecnologia” e “poder”.

As variáveis operacionais serão consideradas a partir das seguintes explicações/definições: para “tecnologia” seguimos a linha do W. Brian Arthur (2009); para “poder” a interpretação adotada é a do Pierre Bourdieu (1976); e já para a “cientificidade”, diante da magnitude da sua complexidade - e das variações de ordem teórico-histórico-geográfica-, consideramos, de forma relativa, as proposições de Karl Popper (1980).

Essas hipóteses foram verificadas, entre 2008 e 2012, a partir de um estudo de caso com 12 projetos do FP7, sendo que, a delimitação das variáveis, no sentido do que era observado em cada projeto, seguiu, na conjugação dos nossos objetivos, as seguintes deliberações: o primeiro ponto era: o projeto (seja ele qual for) “transita por qual(is) estrato(s) da CC (“pesquisa”, “sistema” e/ou a “sociedade”)?”; e, de forma complementar, questionamos: “com base no conteúdo dos relatórios e/ou comunicações do projeto considera-se, ou não, que a acepção tradicionalmente em vigor da CC é condizente e/ou suficiente para explicar o fenômeno?”; terceiro aspecto de análise gira em torno das três categorias elencadas na hipótese subjacente: a “tecnologia”, o “poder” e a “cientificidade”. Obviamente cada um desses elementos nos levaria a “n” interpretações e análises por distintas perspectivas que seriam infundáveis e insustentáveis em profundidade, em função do tempo de execução de uma pesquisa. Nesse sentido, interessava, aqui, num primeiro momento, detectar se um determinado projeto (dos 12 em questão), em algum dos seus relatórios e/ou comunicações, discutia ou considerava como relevante uma ou mais dessas três categorias supracitadas e, se assim o fosse, numa segunda instância, medir, via as relações “positivas” e “negativas” entre as variáveis, quais eram as interlocuções e proposições lançadas para e entre tais elementos confirmando, ou não, nossas hipóteses.

Pesquisa de Base com enfoque nos Projetos FP7 da União Europeia

A pesquisa de Base iniciou-se em 2008 e foi finalizada em 2012 (Gomes, 2012) e os dozes projetos analisados podem ser assim sintetizados: o LiquidPub articula o ciclo da CC como um todo e propõe, dentre outras coisas, alternativas para as publicações científicas, o Driver II trabalha na perspectiva dos repositórios e das “Enhanced Publications”, o Wf4ever com a preservação a longo prazo dos “scientific workflows”, o Tina com a paisagem científica (cientometria) e com contribuições na ordem da força

entrelaçada ao “poder”, o Gsdp com o diálogo político e o Soap com a preocupação econômica, o Kiwi com a *wiki* + web semântica, o Papyrus com os conteúdos transpostos de uma disciplina para a outra; o Sisob com a apropriação social do conhecimento, o Casc com o envolvimento das pessoas com a ciência, o Macospol na ótica do mapeamento de controvérsias e o Bison com a conjugação do raciocínio em rede + o pensamento criativo.

Os resultados parciais da pesquisa são abaixo apresentados:

SÍNTESE DOS RESULTADOS:

(1) O projeto transita por qual(is) estrato(s) da CC (“pesquisa”, “sistema” e/ou a “sociedade”)?

Os dados oriundos da leitura e análise qualitativa dos relatórios/comunicações, conjugados com os aspectos quantitativos da manifestação das expressões (“unidades de registro”), originou a seguinte relação dos projetos por segmento da acepção tradicionalmente em vigor da CC:

Tabela 1: Projeto por segmento da acepção tradicionalmente em vigor da CC

Projetos	Pesquisa	Sistema	Sociedade
Bison			
Casc			
Driver II			
Gsdp			
Kiwi			
LiquidPub			
Macospol			
Papyrus			
Sisob			
Soap			
Tina			
Wf4ever			

Obtivemos, portanto, 8 resultados para “pesquisa”, 9 para “sistema” e 8 para “sociedade”, todas as categorias, logo, com mais de 50% de ocorrências no total dos 12 projetos da amostra. A inferência para assinalar um “quadrado” e não outro foi realizada a partir de dados complementares que possibilitaram, por aspectos diferentes, concluir com precisão qual era a melhor alternativa. No caso, por exemplo, do “sistema”, na ótica do que consideramos na acepção tradicionalmente em vigor (“ou seja, a informação que flui de e entre os intermediários da cadeia de comunicação - Meadows, 1974 - editoras,

bibliotecas, agências de financiamento, dentre outros”), muitas vezes, na “unidade de registro” não aparecia a expressão em si, mas, com a leitura, era factual que o projeto estava lidando com os processos da CC, mesmo sem intitulá-los assim e vice-versa. Convém salientar, também, que o nosso *corpus* documental, mesmo não sendo quantitativamente equilibrado (para cada projeto tínhamos um número diferente de documentos sendo analisado⁶) - e nem era essa a intenção, já que contávamos com projetos já finalizados e outros em andamento - era bem representativo, no âmbito do FP7, da realidade da CC na contemporaneidade. Para além do *corpus* documental, nossos desfechos foram rematados por outras informações circunstâncias - a exemplo dos vídeos, no caso do Tina.

(2) Com base no conteúdo dos relatórios e/ou comunicações do projeto considera-se, ou não, que a acepção tradicionalmente em vigor da CC é condizente e/ou suficiente para explicar o fenômeno?

No que tange à questão acima, a resposta foi positiva para todos os projetos (100%), isto é, não detectamos outras definições para a comunicação científica.

(3) O projeto articula a "tecnologia" com o "poder" e a "cientificidade" nos moldes da nossa hipótese subjacente?

(3a) E minimiza os aspectos atrelados ao Poder?

(3b) Por outro lado, garante a Cientificidade?

Dos 12 projetos, somente 3 (25%) (LiquidPub; Tina; e Sisob) respondem positivamente para a pergunta “o projeto articula a ‘tecnologia’ com o ‘poder’ e a ‘cientificidade’ nos moldes da nossa hipótese subjacente?”, sendo que, desses, um deles (o LiquidPub) não se coaduna de forma plena - não garantindo a “cientificidade” / só complementando via a “atribuição de créditos com base em redes sociais, etc”.

Em seguida temos, pois, a apresentação de dois “suplementos”. Como a própria nomenclatura já designa, o “suplemento” como “o que se dá a mais” visou ampliar o entendimento da realidade existente na CC como forma de compreendermos como a *práxis* do fenômeno se articula na contemporaneidade. No primeiro caso, um “quadro-esboço” foi elaborado com o propósito de facilitar a compreensão do leitor sobre os campos de inserção dos projetos a partir da hipótese subjacente da pesquisa-base, no entanto, como é patente, essa representação gráfica não busca exprimir, rigorosamente, toda a complexidade da CC - seria, ou é, tecnicamente difícil comprimir todos os processos e ações da CC em um único quadro. Como é pertinente salientar, cada uma das nove intersecções oriundas das 6 variáveis (*pesquisa, sistema e sociedade / tecnologia, poder e cientificidade*) não foram preenchidas de antemão porque, indiscutivelmente, não partimos de “categorias pré-definidas” estando, por conseguinte,

⁶ Do conjunto dos relatórios, o do Kiwi foi o mais complexo de análise em função do suporte técnico e, na oposição deste, figura-se o Soap.

de forma exploratória, isentos de um “quadro preestabelecido”. Ao mesmo tempo, a pesquisa-base, opera no nível “macro”, quer dizer, “holístico” e, nesse sentido, também não direcionamos os esforços para os componentes do microambiente como, por exemplo, tudo que comportaria a interface “poder” *versus* “pesquisa” e assim sucessivamente. Já no segundo caso, o suplemento era alimentado pelas 10 palavras (com exceção das preposições e números) mais citadas no conjunto de relatórios/comunicações por projeto analisado visando, pois, o melhor enquadramento destes nas questões lançadas e, conseqüentemente, maior sinergia com os objetivos da presente investigação. Os suplementos são, portanto, deste modo sistematizados:

SUPLEMENTOS

A) Campos de inserção do projeto (no âmbito da hipótese subjacente):

Agrupando todos os dados das tabelas expostas após a análise de cada um dos projetos do FP7 temos, pois, o seguinte quadro-síntese de ocorrência de cada uma das variáveis por projetos:

Tabela 2: Quadro-síntese de ocorrência das variáveis por projetos FP7 analisados

Cientificidade	Soap	Driver II LiquidPub Sisob Soap Tina Wf4ever	Soap
Poder	LiquidPub Soap	Gsdp LiquidPub Sisob Soap Tina	Soap
Tecnologia	Driver II Gsdp Kiwi LiquidPub Papyrus Sisob Wf4ever	Bison Driver II Gsdp Kiwi LiquidPub Macospol Sisob Tina wf4ever	Driver II Kiwi LiquidPub Macospol Sisob Wf4ever
	Pesquisa	Sistema	Sociedade

Tal quadro-síntese, somado às observações detalhadas no “quadro-resumo de dados e análise” (Gomes, 2012) de cada projeto, nos remete, pois, aos seguintes resultados: a maioria (9 de 12, ou seja, 75%) dos projetos está concentra da interface “tecnologia” *versus* “sistema” e somente um (Soap) nas conjugações “pesquisa” *versus* “cientificidade”, “sociedade” *versus* “cientificidade” e “poder” *versus* “sociedade”. 11 projetos, com exceção do Casc, cujo quadro ficou em branco, estão em, pelo menos,

uma linha da acepção tradicionalmente em vigor da CC (pela ótica da “pesquisa” e/ou “sistema” e/ou “sociedade”) em interface com, pelo menos, também, outra camada (“tecnologia”, “poder” e/ou “cientificidade”).

B) Frequência de Palavras:

Em relação a frequência de palavras por projeto foi encontrado:

Tabela 3: Palavras mencionadas por projeto

Projetos	Palavras (e derivações) associadas com as variáveis - do objetivo e da hipótese	Palavras relacionadas com o conteúdo do projeto	O nome próprio do projeto	Outras
Bison	Systems	Information; Data; Network; Discovery; sources	x	Different; New; Gene; Vertices
Casc	Scientific; Society	Science; Public; People; Initiatives; Engagement	Casc	Project; Tools
Driver II	Research	Repositories; Repository; Digital; Access; Open Metadata; European	Driver	Page
Gsdp	Systems; System; Social	Risk; Global; Models; Policy; Science; Information	x	Different
Kiwi	System	Knowledge; User; Content; Data; Tag; Information; Software	Kiwi	Project
LiquidPub	Research	Reputation; Review; Web; Information; Model; Sko; Authors	LiquidPub	Use
Macospol	x	Controversies; Mapping; Controversy	Macospol	Case; Study; Students; Course; Realised; Work
Papyrus	x	News; Ontology; Content; Search; History	Papyrus	User; Project; Material
Sisob	Science; Social	Data; Knowledge; Review	Sisob	Network; Networks; Mobility; Page
Soap	Society	Journals; Open; Publishers; Access; Journal; Articles; Publisher; Publishing; Doaj; Article	x	X
Tina	x	Network; Map; Maps; Semantic; Portfolio	x	Terms; Project; Find; User; Queries
Wf4ever	Research	Data; Workflow; Information; Objects	Wf4ever	Requirements; User; Project;

			Provenance
<p>Os resultados acima, sem nenhuma pretensão de alcançar, analisar ou desenvolver nenhum método bibliométrico, buscou estampar, via a frequência de palavras, com o auxílio do NVivo9, o conteúdo dos projetos somando a análise qualitativa oriunda da leitura dos documentos.</p>			

Nosso objetivo específico foi alcançado (“verificar se os projetos do FP7, objetos do estudo de caso, transitam somente pelos estratos da ‘pesquisa’, do ‘sistema’ e/ou da ‘sociedade’”) e as hipóteses não foram confirmadas pela pesquisa empírica - a noção de comunicação científica considerada a partir da acepção tradicionalmente relacionada ao tripé “pesquisa, sistema e sociedade” é condizente e suficiente para explicar o fenômeno na contemporaneidade e a hipótese subjacente não é sustentável - nove projetos do FP7, ou seja, 75% do total não legitimam a mesma.

Possível Re-escritura da Comunicação Científica

A CC é composta por diversos micro e macroambientes e, em todos eles, encontramos múltiplas forças que exercem uma pressão sobre o sistema. Tais forças podem atuar tanto de forma “isolada”, como *input* ou *output* de diversas naturezas, ou em “sintonia” uma para com a outra, visto que, entre as mesmas também existem influências mútuas e peculiaridades. E, ainda, por se tratar de um fenômeno em constante mutação, as interações entre o sistema de CC e as forças não são “estanques”, ou seja, por mais que possamos conhecer todas as variáveis de preponderância, a ocorrência dos fatos, no tempo e no espaço, nunca é “repetida”. Caímos, pois, no paradoxal da necessidade da compreensão da CC pelo viés “holístico” e na dificuldade de generalizações que possam ser aplicadas às mais díspares sociedades. Cientes dessa conjuntura, considera-se válido o desafio em questão e elencam-se as principais forças do macroambiente, já que, optou-se, desde o início, por uma análise nesse nível de perspectiva e não pelo viés micro, que atuam como elementos volúveis no sistema de CC e que foram identificadas quando da constituição da pesquisa de base (Gomes, 2012), a saber: a “disciplinaridade”, “economia”, “pessoas/cultura”, “poder”, “política” e “tecnologia”.

Nomeadamente, o *fluxo* da CC é atrelado aos “processos” e estes podem ser observados, por exemplo, através do diagnóstico proposto por Bjork (2007) e outros e inerente ao sistema da CC estão seus *atores sociais*, cada qual com um contínuo de problemáticas constantes e/ou mutáveis a partir da variável temporal já descrita, muitas vezes, nos modelos tradicionais de representação da CC como, no caso, do de Garvey e Griffith (1979). O pilar da CC é a “cientificidade” - assegurada pelo *peer review* - e acoplada a esta estão os processos e fluxos que são influenciados pelas forças do macroambiente aqui compostas a partir das 6 (seis) grandes categorias, destacadas acima, e que estão sempre inter-relacionadas. Tudo isto envolto ao que consideramos ser

o “macroambiente tecnológico” que diretamente conjuga-se com a contemporaneidade e se sobrepõe aos demais macroambientes.

O panorama traçado não é, pois, fechado em suas nuances, a “tecnologia”, por exemplo, é uma força e também compõe o “macroambiente tecnológico”, assim como, a “cientificidade” pode ser considerada uma das variáveis pela ótica da nossa hipótese subjacente e é, concomitantemente, por tudo que a pesquisa de base (Gomes, 2012) revelou, o “pilar” da CC.

Sinteticamente temos, portanto, os seguintes atributos que podem complementar (não substituir), pelo viés holístico, a noção de comunicação científica considerada a partir de sua acepção tradicionalmente relacionada ao tripé “pesquisa, sistema e sociedade”:

Tabela 4: Atributos que podem complementar, pelo viés holístico, a noção de comunicação científica considerada a partir de sua acepção tradicionalmente relacionada ao tripé “pesquisa, sistema e sociedade”

MACROAMBIENTE TECNOLÓGICO
+
CONJUNTO DE FORÇAS: DISCIPLINARIDADE, ECONOMIA, PESSOAS/CULTURA, PODER, POLÍTICA E TECNOLOGIA
+
ATORES SOCIAIS - E SUAS PROBLEMÁTICAS - NA VARIÁVEL TEMPORAL
+
PILAR = CIENTIFICIDADE - assegurada pelo <i>peer review</i>

Além desses atributos, incluiremos agora, como uma espécie de prognóstico da CC, um cogito sobre uma nova maneira de pensar a mesma a partir de uma possível “*re-escritura*”. Adotamos esse termo, que nos parece preferencial, a outras expressões possíveis como, por exemplo, “pós-comunicação científica”, tendo como referência as considerações lançadas por Lyotard (1989) em um ensaio no qual o filósofo francês trata de algumas questões ligadas ao uso do prefixo “pós” associado à terminologia “pós-modernidade”:

Este título, reescrever a modernidade [...] Parece-me bastante preferível às rubricas habituais como “pós-modernidade”, “pós-modernismo”, “pós-moderno”, sob as quais é geralmente colocado este tipo de reflexão. A vantagem consiste em duas deslocções, a transformação do prefixo “pós” em “re”, do ponto de vista léxico e a aplicação sintática do prefixo assim modificado no verbo “escrever” em vez do substantivo “modernidade”.

Esta deslocção dupla indica duas direções principais. Primeiramente faz realçar a futilidade de qualquer periodização da história cultural em termos de “pré” e de “pós”, de antes e de depois pelo simples fato de não resolver a posição do “agora”, do presente a partir do qual é suposto podermos adotar uma perspectiva legítima sobre um decurso cronológico.

[...] nem a modernidade nem a dita pós-modernidade podem ser identificadas e

definidas como entidades históricas claramente circunscritas, onde a segunda chegaria sempre “depois” da primeira. Falta precisar, pelo contrário, que o pós-moderno está já compreendido no moderno pelo fato de que a modernidade, a temporalidade moderna comporta em si o impulso para se exceder num estado que não é o seu. E não apenas a exceder-se nele mas a converter-se nele como uma espécie de estabilidade última como seja a que visa por exemplo o projeto utópico, mas também o simples projeto político presente nos grandes elogios da emancipação. Devido à sua constituição, e sem descanso, a modernidade está grávida do seu pós-modernismo.

[...] É possível agora clarificar uma segunda acepção, diferente, deste “re”. Ligado de maneira fundamental à escrita, ele não significa de maneira nenhuma um retorno ao começo mas, de preferência aquilo que Freud designou por “perlaboração”, a “Durcharbeitung”, ou seja, um trabalho dedicado a pensar no que, do acontecimento e do sentido de acontecimento, nos é escondido de forma constitutiva, não apenas pelo pressuposto anterior, mas também por estas dimensões do futuro que são o pro-jeto, o pro-grama, a prospectiva, e mesmo a pro-posição [...] (Lyotard, 1989, p.35).

Posto isto, os princípios básicos que poderão nortear essa possível “re-escritura da comunicação científica” envolveriam:

- ✓ De um lado as 3 grandes “composições” da CC noutoras citadas que compõem a acepção tradicionalmente em vigor do fenômeno: a *pesquisa*, o *sistema* e a *sociedade*. Essas instâncias fazem parte da “essência” da CC e poderiam ser designadas como uma espécie de “níveis de continuum”⁷ em relação ao fenômeno.
- ✓ E, de outro, com o deslocamento de conceitos da área de estatística, as “três sigma” - “frequente o suficiente para chamar a atenção, mas não para descartar a possibilidade de que se trate de uma flutuação casual. ‘Com três sigma você tem uma evidência, mas ainda não tem uma descoberta’, diz Shellard”⁸ - compostas por camadas correspondentes: a *tecnologia*, ao *poder* e a *cientificidade*.

E, nos moldes do quadro-resumo de dados e análise da pesquisa de base (Gomes, 2012), é como se a sequência “horizontal” da esquerda para a direita fosse formada pela: *pesquisa*, o *sistema* e a *sociedade* e, de forma “vertical”, abaixo estaria a *tecnologia*, depois o *poder* e por cima a *cientificidade*. Os 6 componentes possuem uma autonomia

⁷ Essa expressão é utilizada em diversas outras áreas como, por exemplo, no contexto biopsicosocial no qual os seis “níveis do continuum” são: biológico, pessoal, relacional, familiar, comunitário e social. Com as devidas adaptações, no campo da CC, quando Bjork (2007), por exemplo, apresenta o seu diagnóstico, o mesmo nada mais é do que “níveis de continuum”, uma vez que, descreve “partes de um todo” relacionadas ao fenômeno em si.

⁸ Fonte: Do Editor de Ciência. *Se partícula existir, LHC deverá detectá-la*. Disponível em: <<http://www1.folha.uol.com.br/fsp/ciencia/fe0704201102.htm>>, acessado em 7 de abril de 2011.

“individual” - que, inclusive, faz agregar outras variáveis tais como os “atributos que complementam, pelo viés holístico, a noção de comunicação científica considerada a partir de sua acepção tradicionalmente relacionada ao tripé ‘pesquisa, sistema e sociedade’” expostos na tabela 4 - e, ao mesmo tempo, sofrem e exercem influências (ou pressões) dos demais, por exemplo, a *pesquisa* interfere no *sistema* que, por sua vez, modifica a *sociedade* e, ao mesmo tempo, a *pesquisa* possui uma camada de *tecnologia*, de *poder* nas relações e de *cientificidade* e assim sucessivamente, numa espécie de miscigenação sempre em efervescência e com múltiplas interdependências e correlações, da teoria com a *práxis* e vice-versa, podendo seguir inúmeras possibilidades de interpretação, visto que, a sociedade contemporânea está alicerçada em “parâmetros” onde já não se tem uma “superação” ou “esgotamento” de uma fase a outra, tudo está imbricado, interligado, justaposto e com a CC não é diferente. A complexidade advém da CC não poder ser resumida em um “objeto” único e palpável, o que simplificaria, e muito, o entendimento do fenômeno e só isto. As teias que podem ser formadas vão se multiplicando e passam por diversos processos de mutação, inclusive, nos próprios processos.

Os 6 componentes (se é que podemos intitulá-los assim) da CC divididos em dois grupos (os *níveis de continuum* e as *três sigma*) são, portanto, elementos existentes, tanto na teoria como no empírico/*práxis* da CC na contemporaneidade, no entanto, o que designamos como uma possível “re-escritura da comunicação científica” parte de um “movimento” entre as instâncias que, por sua vez, não se realiza plenamente em todos os níveis. Em termos objetivos, essa *re-escritura* se concretizaria quando, nos três de *níveis de continuum* (pesquisa, sistema e sociedade), a tecnologia fosse “capaz o suficiente” de minimizar os aspectos negativos atrelados ao *poder* sem, ao mesmo tempo, descaracterizar a CC em termos de *cientificidade*. Numa espécie de “jogo” que, só se “ganha”, quando se “perde” no *poder* e se “mantém” na *cientificidade*.

Apesar da abstração aparente, não seria utopia nenhuma a existência de ferramentas - e o Tina e o Sisob, de alguma maneira são, dos projetos analisados no âmbito da presente pesquisa, os representantes que “chegam lá”, principalmente nos *níveis de continuum* do “sistema” e da “sociedade” - que, via *tecnologia*, minimizassem o *poder* e mantivessem a *cientificidade*. Quando esse mesmo fato ocorrer em todos os *níveis de continuum*, entrelaçados com os desdobramentos que envolvem os atores da CC em cada instância e, assim o esperamos, será, materializadamente inaugurada, essa possível “re-escritura da comunicação científica” para, em seguida, uma ciência efetivamente aberta se consolidar.

A noção de “re-escritura da comunicação científica” em seus pormenores

Esse item a medida que prima por esclarecer os “pormenores” da *re-escritura* esbarra numa dualidade contraditória a partir do momento que concluímos que a CC não pode ser resumida em um objeto único e palpável e, nesse sentido, ser capaz de comportar ações e processos repetitivos e iguais. A questão é: até que ponto exemplificarmos todas as nuances da *re-escritura* da CC não delimitaremos, no sentido

de “fecharmos” em espaços concretos, o que a mesma, por si só, busca romper? Quando afirmamos que “a tecnologia deveria minimizar o *poder* e manter a *cientificidade* em todos os *níveis de continuum*”, objetivamos que a proposição pode ser aplicada em todas as conjunturas, pela ótica dos diferentes atores sociais e em contexto socioculturais e econômicos, o quão mais dispersos melhores, sendo, portanto, complicado (e perigoso) nos sujeitarmos a exemplos deste ou daquele caso. Por ora, pois, esclareceremos algumas circunstâncias básicas da possível *re-escritura* da CC:

Para além das explicações/definições adotadas para cada uma das variáveis da pesquisa (Gomes, 2012) temos que: o “minimizar”, na expressão de base da *re-escritura*, é sinônimo de “diminuir” no sentido “negativo” o que a própria expressão “poder” traz embutida e, mesmo dentro dessa única sigma, em algumas circunstâncias, o “poder” é benigno, como quando atravessamos o nível da “sociedade” - e, sem nenhum tipo de utopia ingênua, muito se ganharia se a sociedade conseguisse ter o controle (ou “poder”, para mantermos a expressão) sobre a divulgação científica⁹ e/ou pudesse, como observamos no Sisob, ampliar o escopo das investigações, em diversas ordens, em prol de um maior impacto das investigações na sociedade e consequentemente uma expansão da qualidade de vida em geral. Por outra via, quando abordamos o sem “subtrair” seu “pilar” principal, ou seja, sua “cientificidade”, estamos nos referindo a manutenção de variáveis consideradas “básicas” ou “estratégicas” para determinar se uma pesquisa traz em seu bojo características de “confiabilidade” e estas, também, se moldam de acordo com a área disciplinar e os contextos históricos, geográficos, etc. Mas, porque propor o “manter” e não o “ampliar”? O “ampliar” seria o conveniente a longo prazo, inclusive em se pensando numa ciência aberta, mas é difícil de “medir” a curto, ou seja, num por vir próximo, “só manter” já seria de grande valia - mesmo com o anseio de que o “ampliar” possa ser o necessário.

Percebemos, também, que a “pesquisa” é dos 3 níveis propostos, a primeira e, ao mesmo tempo, a que menos elementos concretos temos de que a possível *re-escritura* da CC irá efetivar-se (no “sistema” podemos tomar como referência o Tina e na “sociedade” o Sisob). Uma das linhas de interpretação é que o investigador possui, muito mais, a tecnologia como uma ferramenta de apoio quando da descoberta científica, da busca de informações, etc, do que algo potencializador de uma mudança na e entre as pessoas ou processos, visto que, o caráter individual ainda é o soberano em grande parte do cenário científico contemporâneo, sendo as relações, nesse nível, difíceis de serem mensuráveis e, mais ainda, de serem transformadas. O lado oposto é que esse mesmo nível da “pesquisa” é o impulsionador das ações da CC e a “resistência” para com as mudanças, além de “fonte inspiradora” para a possível *re-escritura* que estamos aqui explanando. Enfim, a *re-escritura* da CC substanciará algo concreto quando, conforme relatamos,

⁹ Subvertendo o papel da “divulgação” para o “diálogo” cuja centralidade deixa de ser o pesquisador citamos o “Nordic Network for the Study of the Dialogic Communication of Research” - Disponível em: <<http://dialogue.ruc.dk/>>.

Veja também o texto de Peter Maesele intitulado *Science and technology in a mediatized and democratized society* de 2007.

todos os *níveis de continuum* serem atingidos pela *tecnologia*, minimizando o *poder* e mantendo a *cientificidade*.

Conjuntura atual da pesquisa

Depois de sete anos da pesquisa com os doze projetos do FP7 estamos, agora em 2019, a partir dos mesmos procedimentos metodológicos, analisando outros projetos do Horizon 2020 da UE. Nossa intenção é testar as mesmas hipóteses e verificar a evolução, ou não, das variáveis em torno de uma possível *re-escritura* da comunicação científica na contemporaneidade. Os dados ainda são parciais e não serão expostos aqui mas ficam à disposição para outros entrelaçamentos a partir de possíveis contatos por e-mail com as autoras do presente texto.

Considerações finais

Quando tratamos de uma possível *re-escritura* da comunicação científica estamos, nada mais, nada menos, lidando diretamente com um componente “não atual” da CC que surge enquanto prognóstico do que, para nós, constitui-se o grande contributo desta pesquisa. No entanto, pensar esse “por vir” esbarra em nuances que devem acarretar outros estudos mais aprofundados que possam considerar desde os aspectos que surgem com o “senso comum” (como, por exemplo, “mas por trás de toda tecnologia existe o homem, portanto, é impossível minimizar os aspectos negativos atrelados ao poder” ou, ainda, “mas o poder aí nunca poderá ser medido porque é sempre implícito” ou a “cientificidade será sempre associada ao *peer review*”, etc.) até noções mais complexas de transformação do sistema da CC que contribuam, efetivamente, para “inspirar” uma ciência aberta, igualitária em suas oportunidades e justa nos seus arranjos. Essas “considerações finais” são, portanto, como é perceptível, o “início” de um longo caminho a ser trilhado cujo escopo pode apresentar bifurcações dispersas e curiosas, além de ser, também, como a própria pesquisa em questão, “the view from here”¹⁰...

Referências

- ARTHUR, W. Brian. *The nature of technology: what it is and how it involves*. New York: Free Press, 2009.
- BJORK, B.-C. A Model of Scientific Communication as a global distributed information system. In: *Information Research*, v.12, jan, 2007.
- BOURDIEU, P. Le Champ scientifique. In: *Actes de la Recherche en Sciences Sociales*, n.2/3, jun. 1976.
- GARVEY, W. D. *Communication: the essence of science*. USA: Pergamon Press, 1979.

¹⁰ Expressão utilizada como título do capítulo 9 do livro da Christine L. Borgman (2007) e que, aqui, referendamos.

- _____.; GRIFFITH, B. C. Communication and Information Processing within scientific disciplines: empirical findings for Psychology. In: *Information Storage and Retrieval*, v.8, p.123-136, 1972.
- _____.; GRIFFITH, B. C. Scientific Communication as a social system. In: GARVEY, W. D. *Communication: the Essence of Science*. USA: Pergamon Press, 1979.
- _____ et al. Research studies in patterns of Scientific Communication: I, general description of research program. In: GARVEY, W. D. *Communication: the Essence of Science*. USA: Pergamon Press, 1979.
- GOMES, Cristina Marques. *Comunicação Científica: cartografia e desdobramentos*. São Paulo: ECA/USP, 2012 (Tese de Doutorado).
- _____. *Comunicação Científica: alicerces, transformações e tendências*. Portugal: Livros LabCom, 2013.
- HURD, J. M. Models of Scientific Communications systems. In S. Y. CRAWFORD, J. M. HURD, & A. W. WELLER, *From Print to Electronic: The Transformation of Scientific Communication*. USA: Information Today, 1996.
- _____. The transformation of Scientific Communication: a model for 2020. *Journal of the American Society for Information Science and Technology*, v.51, p.1279-1283, 2000.
- _____. Scientific Communication: new roles and new players. In: *Science and Technology Libraries*, v.25, n.1-2, p.5-22, 2004.
- LYOTARD, Jean-François. *O inumano - considerações sobre o tempo*. Lisboa: Editorial Estampa Ltda, 1989.
- MEADOWS, A.J. *Communication in Science*. London: Butterworths, 1974.
- POPPER, Karl R. *Conjecturas e refutações*. Brasília: Editora da UnB. 1980.

A multidimensionalidade da Ciência Aberta

Cristina Marques Gomes¹, Adriana Carla S. Oliveira²

¹ Departamento de Ciências da Comunicação - Universidade Federal de Santa Maria (UFSM). Email: cristina@alumni.usp.br

² Professora Colaboradora da Universidade Federal do Rio Grande do Norte e da Escola de Magistratura do Rio Grande do Norte. Email: adrianacarla.a@gmail.com

Resumo

O presente artigo apresenta o conceito de Ciência Aberta em uma perspectiva multidimensional. Parte do pressuposto de que a Ciência Aberta é um conceito guarda-chuva constituída por dimensões. Foram consideradas no estudo as dimensões: epistemológica, teórica, política, morfológica, técnica e ética. O trabalho atende ao eixo temático Ciência Aberta e dados abertos: promoção, práticas e políticas, visibilidade da ciência, éticas e direitos autorais. Com base na temática levanta-se a seguinte questão de pesquisa: As dimensões propostas no estudo configuram o conceito guarda-chuva da Ciência Aberta das atuais práticas científicas? O estudo tem como objetivo geral, descrever as dimensões da Ciência Aberta, e como objetivos específicos, contextualizar a Ciência Aberta e a *e-Science* no cenário internacional, compreender o escopo dos movimentos *Open Access* e da *Open Science* que figura na atual ciência contemporânea e, representar por meio de um modelo conceitual as dimensões da Ciência Aberta. O método científico adotado partiu do estudo das dimensões de pesquisa de Bufrem, que tem como premissa epistemológica o método quadripolar. A pesquisa é puramente qualitativa e adotou a análise de conteúdo de Bardin e o *Software NVivo* para apoiar a categorização de conteúdo. Concluiu-se que a abordagem multidimensional é norteadora para uma política orientada à Ciência Aberta e aos Dados Científicos, bem como o direcionamento à *práxis* da *e-Science*. Essa abordagem representa a multidimensionalidade da ciência contemporânea para o contexto nacional.

Palavras-chave: ciência aberta, *e-Science*, dados científicos, multidimensionalidade.

Abstract

This paper presents the concept of Open Science in a multidimensional perspective. It starts from the assumption that Open Science is an umbrella concept made up of dimensions. They were included in the context as dimensions: epistemological, theoretical, political, morphological, technical and ethical. Open source Thematic work Open Science and are published: promotes, news and policies, visibility of science, ethics and copyright. Based on your question, the following research question arises: How do the proposed questions not form the concept of open science? The objective of the study is to describe the scope of open and open access as the scope of Open Science, and as specific objectives, to contextualize

Open Science and Science in no international scenario. By means of a conceptual model like the dimensions of Open Science. The thematic method adopted is the method of the research scripts of Bufrem, whose premise is the epistemological focus of the quadrupolar method. The research is purely qualitative and adoptived with content analysis of NVivo software and software to support a categorization of content. It concludes that it is a multidimensional approach to a science-oriented policy Openness and Scientific Data, as well as the direction to the praxis of e-Science. This approach represents a multidimensionality of contemporary science to the national context.

Keywords: open science, e-Science, Scientific data, multidimensionality.

Introdução

O texto completo partirá do resumo supracitado apresentando os problemas das pesquisas, o percurso metodológico empregado, os resultados obtidos e a correlação desses com a Ciência Aberta numa perspectiva multidimensional ao que se tem discutido, em âmbito internacional, sobre o tema.

O fio condutor da pesquisa tem como ponto de partida o entendimento das diferentes nuances do modelo vigente da comunicação científica – ênfase na publicação de resultados científicos – e do emergente nomeado *e-Science* – com foco no uso e reuso de dados abertos de pesquisa. Os processos e práticas da comunicação científica, que por séculos, foram apoiados pela cultura impressa que tornavam públicos apenas os resultados de investigação, foram repensados a partir de movimentos emergentes, tais como: *Open Access*, *Open Science*, *Fourth Paradigm* e *Data Science*.

Tais movimentos considerados pressupostos das mudanças no *modus operandi* de gerenciar, tratar, disponibilizar, acessar, usar e reutilizar as publicações científicas impressas e eletrônicas, conduzem também para o gerenciamento e a curadoria dos dados científicos abertos, como produto primário de pesquisa. Práticas científicas concentram esforços entre diferentes colaboradores e ambientes com ativistas, pesquisadores e instituições de fomento, acadêmicas, científicas e governamentais, e são direcionadas à promoção da investigação científica colaborativa, aberta e inovadora.

Com base na temática levanta-se a seguinte questão de pesquisa: as dimensões propostas no estudo configuram o conceito guarda-chuva da Ciência Aberta das atuais práticas científicas? Em observância à referida questão, a hipótese defendida, é a de que sendo a Ciência Aberta um conceito guarda-chuva, a sua *práxis* se constitui em diferentes dimensões e apregoa o novo fazer científico contemporâneo.

O estudo tem como objetivo geral, descrever as dimensões que configuram a *práxis* da Ciência Aberta, e como objetivos específicos, contextualizar as dinâmicas contemporâneas científicas da Ciência Aberta e da *e-Science* no cenário internacional, compreender o escopo político-normativo-legal dos movimentos do *Open Access* e da *Open Science* e, representar por meio de um modelo conceitual

as dimensões da Ciência Aberta.

Sob o ponto de vista dos procedimentos metodológicos, a pesquisa é predominantemente qualitativa e adotou a análise de conteúdo de Bardin para a categorização, codificação e inferências do *corpus* de análise. O método científico adotado partiu do estudo das dimensões de pesquisa de Bufrem (2013), que tem como base epistemológica o método quadripolar idealizado por Paul de Bruyne, J. Herman e M. de Schoutheete, e que foi ampliado em duas novas dimensões.

Por fim, a trajetória metodológica tem o objeto de estudo centrado na multidimensionalidade, a partir de domínios específicos e suas relações com objetos distintos do conhecimento. A contribuição de cada domínio específico e sua relação com outros conhecimentos se direciona para o objeto de estudo como recorte multidisciplinar.

A ideia da visão multidimensional parte do *constructo* do estudo que ilustra aspectos políticos, normativos, epistemológicos, teóricos, morfológicos, legais, éticos e culturais, que compõem o atual ambiente colaborativo da prática científica.

O modelo multidimensional traz uma ampliação dos aspectos e características da pesquisa moderna e foi considerado o modelo conceitual. É um modelo norteador para a construção teórico-prática que reflete as novas dinâmicas científicas.

Ciência aberta e a *práxis* da *E-science*

A distinção conceitual entre a Ciência Aberta e a *e-Science* se fez necessária neste estudo com vistas a evidenciar a visão multidimensional que envolve esse contexto emergente. Bem com, explicitar a abordagem conceitual que segue na direção de que a Ciência Aberta é o ambiente inovador, colaborativo, compartilhado, aberto e digital da ciência moderna e a *e-Science* é a representação de uma nova configuração para as práticas científicas.

A priori, entende-se que a Ciência Aberta é uma abordagem para investigações no contexto da comunicação científica e da ciência-cidadã. É um termo derivado da expressão *Open Science* que surgiu no século XVI após a revolução pós-renascentista, com a proprietização dos bens intelectuais, a partir do surgimento da imprensa. Ressurgindo nas últimas décadas como um grande esquema emergido do movimento de acesso aberto (FROSIO, 2014). Apresenta como premissa um modelo que busca o equilíbrio entre a propriedade privada e o acesso aberto do capital intelectual (OLIVEIRA, 2016).

Acerca da Ciência Aberta pode-se afirmar que se constitui como um termo guarda-chuva, amplo e multidimensional, que inclui acesso livre as “publicações científicas, dados científicos abertos, ferramentas científicas abertas, *hardware* científico aberto, cadernos científicos abertos, *wikipesquisa*, ciência cidadã e educação aberta” (ALBAGLI, 2015, p. 15).

É uma ciência voltada para grandes conjuntos de dados que são gerados em sofisticadas bases de dados e *software* que agrega uma ampla gama de simulações, metodologias e modelos efetivados em diversas áreas do conhecimento.

Ademais, a ciência contemporânea é uma prática científica em que outros colaboram e contribuem sempre que os dados de pesquisa estejam disponíveis. Engloba, ainda, ferramentas e recursos tecnológicos digitais e colaborativos,

mecanismos alternativos para a propriedade intelectual, como o uso de licenças *copyleft*, que promovem a reutilização dos resultados de pesquisas (DELFANTI; PITRELLI, 2015) enfatizando o novo fazer científico.

O novo fazer científico pressupõe que regras e práticas sejam ensejadas pela comunidade científica, agências de fomento, governo e instituições de pesquisa e ensino. Esse fenômeno demanda que se pense amplamente nas dimensões epistemológica, política, ético-legal-cultural, morfológica e tecnológica que sustentem a efetivação da ciência aberta.

Além das questões conceituais e metodológicas tradicionais, as práticas científicas contemporâneas exigem uma infraestrutura técnica e tecnológica para apoiar o fazer científico no escopo da *e-Science*.

Acerca do termo *e-Science*, também conhecido como *E-Science* ou *eScience* é definido como a ciência de computação intensiva, realizada em ambientes de rede altamente distribuídos que utilizam volumosos conjuntos de dados, os quais permitem a coleta, processamento, preservação, análise e armazenamento de grande quantidade de dados em diferentes disciplinas. Representa, pois, as práticas preconizadas em um novo paradigma denominado quarto paradigma científico.

A ciência contemporânea do Século XXI, é a pesquisa colaborativa que produz grande volume de dados deu origem aos termos: *Big Data*, *cyberscience*, *cyber infrastructure* e *e-Science*” (COSTA; CUNHA, 2015, p. 128). As inovações, descobertas e teorias sempre fizeram parte do escopo da ciência, contudo, as iniciativas de *e-Science* privilegiam o desenvolvimento de uma ciência aberta (*open Science*), ligada principalmente à disponibilização e manutenção de bases de dados abertas, de acesso público, que subsidiem o trabalho de pesquisa tanto no âmbito individual como no colaborativo.

A formação desses conjuntos de dados é, em geral, decorrente de resultados de projetos de pesquisa financiados com recursos públicos, sendo disponibilizados para acesso livre após um período de carência, ficando disponíveis para outros pesquisadores (APPEL, 2014, p. 15).

Mudanças significativas em ambientes que envolvem disciplinas das ciências naturais e da terra, biotecnologia, medicina genética, biotecnologia e o avanço cada vez mais especializado das engenharias, tecnologias, a exemplo da nanotecnologia, engenharia genética e inteligência artificial estão cada vez mais evidentes.

Além de que, “a relevância dos dados no contexto das “big Sciences”, como Astronomia, Física e Biologia, conduziu não somente ao surgimento de novos modelos de ciência - coletivamente chamados de “quarto paradigma científico” ou “eScience” [...]” (BORGMAN, 2010, p. 2, tradução nossa), mas a novos modelos e práticas a partir da utilização e do valor que os dados de pesquisa passaram a ter em áreas consideradas de alta complexidade e especializadas, dessas chamadas grandes ciências.

A natureza colaborativa e multidisciplinar corrobora como o novo fazer científico através do gerenciamento de dados de pesquisa. A prática científica pressupõe uma infraestrutura que envolva aspectos políticos, institucionais, gerenciais, metodológicos e tecnológicos. No cenário internacional, os esforços para uma ciência orientada a pesquisa aberta e aos dados científicos são apoiados por políticas, diretrizes, modelos, tecnologias e boas práticas.

No Brasil, tal prática está emergindo e se fortalecendo em universidades, institutos de pesquisa, agências de fomento e no governo federal.

Pesquisa aberta no quarto paradigma científico

A evolução da ciência é pautada por revoluções científicas que representam suas práticas e provocam mudanças no *status quo* e no *modus operandi* do pensar e fazer científico. Na contemporaneidade, a ciência e a tecnologia interagem com a comunicação científica e “têm sido responsáveis pelas grandes transformações pelas quais a humanidade passa, alterando ainda de forma desigual, o cotidiano das pessoas” (GASQUE, 2008, p. 30). Tal interação é primordial para o avanço científico e social. Os meios tecnológicos dão suporte às transformações científicas que estão cada vez mais dinâmicas e velozes.

A essa evolução na ciência, por vezes, denominamos de mudança de paradigma. Tal mudança é impulsionado por esforços de cientistas que trabalham incessantemente para que práticas científicas evoluam e novas teorias, dinâmicas e modos de fazer e comunicar a ciência possam surgir, despontando para a mudança de paradigmas.

Mesmo não sendo o precursor, foi o filósofo Thomas Kuhn, o primeiro a adotar o termo paradigma para o contexto da ciência, defendendo que a noção de paradigma é reforçada pelos dogmas, práticas arraigadas e teorias antigas que se mudam ou alteram de acordo com a evolução histórica, científica, econômica, social e cultural na sociedade. Nas práticas científicas alterações e mudanças de paradigmas são representadas por melhorias, ressignificados de teorias, modelos, exemplos e no *modus operandi* existentes.

O pressuposto de Kuhn se deu na tentativa de compreender as divergências e manifestos ocorridos nas práticas científicas. A sua inquietação pairava nas soluções e modelagens conduzidas nas diversas ciências. Afirmou que, “particularmente, fiquei impressionado com o número e extensão das divergências manifestas entre os cientistas sociais sobre a natureza dos problemas científicos legítimos e métodos” (KUHN, 1970, p. viii, tradução nossa).

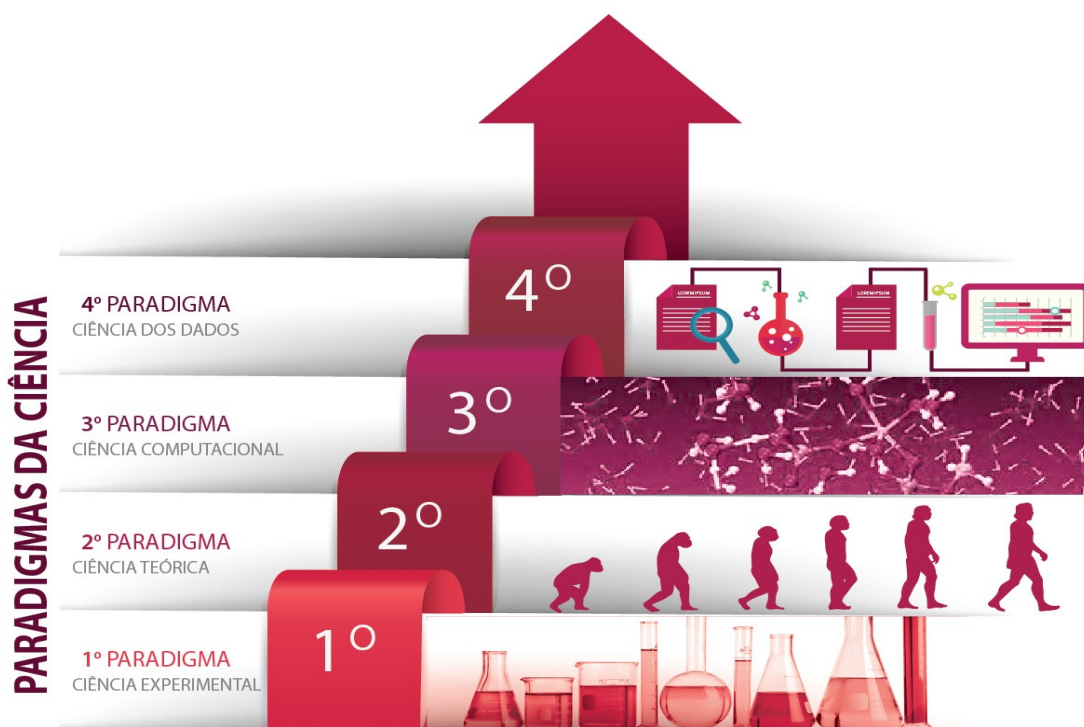
Shuttleworth (2008, *on-line*, tradução nossa) afirma que a mudança de paradigma sugerida por Thomas Kuhn, “não é apenas uma pequena mudança na ciência ou a modificação de uma teoria, mas uma revolução científica e muda completamente o modo como a ciência se configura no mundo [...]”.

Um paradigma pode absorver alguns erros, mas eles eventualmente tornam-se intransponíveis e resultam em uma transição para novas formas de pensar e fazer a ciência. Os paradigmas distinguem-se entre pelos contextos culturais, sociais e históricos da sociedade influenciando a natureza científica.

A ciência contemporânea está experimentando um novo paradigma. O cientista computacional Jim Gray foi o precursor do quarto paradigma científico (*Fourth Paradigm*). A afirmação de Gray (2007) conduz para o entendimento de que o quarto paradigma reflete o atual *status* da ciência aberta e inovadora baseada no uso intensivo de dados de pesquisa para a exploração científica. No entendimento de Jim Gray (2007) o paradigma atual reforça a exploração intensiva de dados unificando os paradigmas anteriores (experiência, teoria e simulação computacional)

em uma nova prática científica. Os paradigmas da ciência representam a evolução histórica e científica dos anteriores paradigmas conforme demonstrado na Figura 1.

Figura 1: Paradigmas da ciência



Fonte: OLIVEIRA (2016).

O primeiro paradigma da ciência está baseado na experimentação. Ocorrido há mais de 1000 (mil) anos com a ciência experimental, teve origem na Grécia antiga e China, onde os filósofos e estudiosos tentavam explicar as suas observações através de leis naturais, ao invés de causas sobrenaturais.

O segundo paradigma baseou-se na ciência teórica com o surgimento de modelos e generalizações. Ocorrido por volta do século XVII, cientistas como Isaac Newton tentaram fazer previsões para os novos fenômenos e verificação de novas hipóteses através da realização de experimentos, originando novas teorias.

O terceiro paradigma, refletido nas últimas décadas, é representado na ciência computacional com base na simulação de fenômenos complexos. Com o advento dos computadores de alto desempenho na segunda metade do século XX, o ganhador do Prêmio Nobel, Ken Wilson, identificou a computação e a simulação como um terceiro paradigma para a exploração científica, tais como: modelagens climáticas ou a formação de galáxias, bem como resolver numericamente os sistemas de equações em grande escala e de forma detalhada (HEY, 2010).

Quanto ao quarto paradigma aflora a geração intensiva de uma grande quantidade de dados que emergem do processo de investigação científica. A mudança está na utilização dos dados primários adjacentes de diferentes disciplinas e utilizados como insumos na investigação científica.

Afirmam Tenopir e outros (2011, p. 1, tradução nossa) que “segundo os paradigmas de investigação anteriores (experimentais, teóricos e computacionais), esta nova era tem sido chamada de ‘o quarto paradigma: descoberta científica do uso intensivo de dados’”. É a junção dos paradigmas anteriores para apoiar o processamento dos dados intensivos na era digital. Complementam ainda que toda a literatura científica está *on-line*, todos os dados da ciência estão *on-line*, e eles interoperam uns com os outros.

Os dados digitais não são apenas as saídas de investigação, mas fornecem subsídios para novas hipóteses, permitindo novos conhecimentos científicos e promovendo a inovação (TENOPIR et al., 2011, p. 1, tradução nossa). Então, além do uso dos dados como insumos primários na pesquisa e a incorporação das tecnologias e de práticas consolidadas nos paradigmas anteriores estão sendo aprimoradas e adicionadas a nova era digital.

Portanto, “o quarto paradigma não está tentando substituir cientistas ou as outras três metodologias, mas exige um conjunto diferente de habilidades” (HEY, 2010, *on-line*, tradução nossa). Sem a capacidade de aproveitar as ferramentas sofisticadas de computadores que manipulam dados, mesmo o perito mais altamente treinado nunca conseguiria descobrir os *insights* que agora estão começando a emergirem. As ferramentas utilizadas para analisar, visualizar, minerar e manipular os dados científicos sistematicamente buscam soluções para desafios globais (HEY, 2010).

A interseção entre os paradigmas reforça que a ciência está sempre em um processo evolutivo e dinâmico. Aprimora e modifica o *status quo* e o *modus operandi* do fazer científico na cadeia científica. De certa forma, o quarto paradigma preconizado por Gray proporciona uma estrutura integrativa que permite os três primeiros a interagirem e reforçam um ao outro, “a exemplo do ciclo científico tradicional no qual a teoria ofereceu previsões que poderiam ser experimentalmente testadas, e esses experimentos identificaram fenômenos que exigiam explanação teórica” (DLYNCH, 2009, p. 177, tradução nossa).

Nessa direção, a ciência aberta mobiliza múltiplos níveis e escopos de abertura, remetendo tanto a um sentido pragmático, de “permitir maior dinamismo às atividades de ciência, tecnologia e inovação, quanto a um sentido democrático, de possibilitar maior diversidade de perspectivas no âmbito da própria ciência e, ainda, maior participação e intervenção da sociedade” (ALBAGLI, 2015, p. 17, grifo nosso). Envolve instituições governamentais, educacionais, científicas e a sociedade em geral.

A tecnologia cada vez mais presente nos processos científicos,

A atividade de pesquisa científica do século XXI produz uma quantidade extraordinária de dados, principalmente em formatos digitais [...]. Isto acontece essencialmente porque a tecnologia digital se torna cada vez mais um elemento onipresente nos processos da construção do conhecimento científico, seja por aumentar a capacidade dos instrumentos científicos, seja por reconstruir realidades por meio de simulação, ou ainda inaugurando formas inéditas de colaboração e compartilhamento de dados e informações (SAYÃO; SALES, 2013, p. 2).

Este cenário, traz à tona discussões sobre a mudança na geração e

disponibilidade de publicações científicas e eletrônicas, conectadas pelas tecnologias digitais de informação e comunicação através da *internet*, abrindo possibilidades de compartilhamento e colaboração coletiva a partir de dados de pesquisa. O quarto paradigma científico apresenta um aprimoramento de métodos, procedimentos e tecnologias com possibilidades de avanços e acessibilidade à pesquisa.

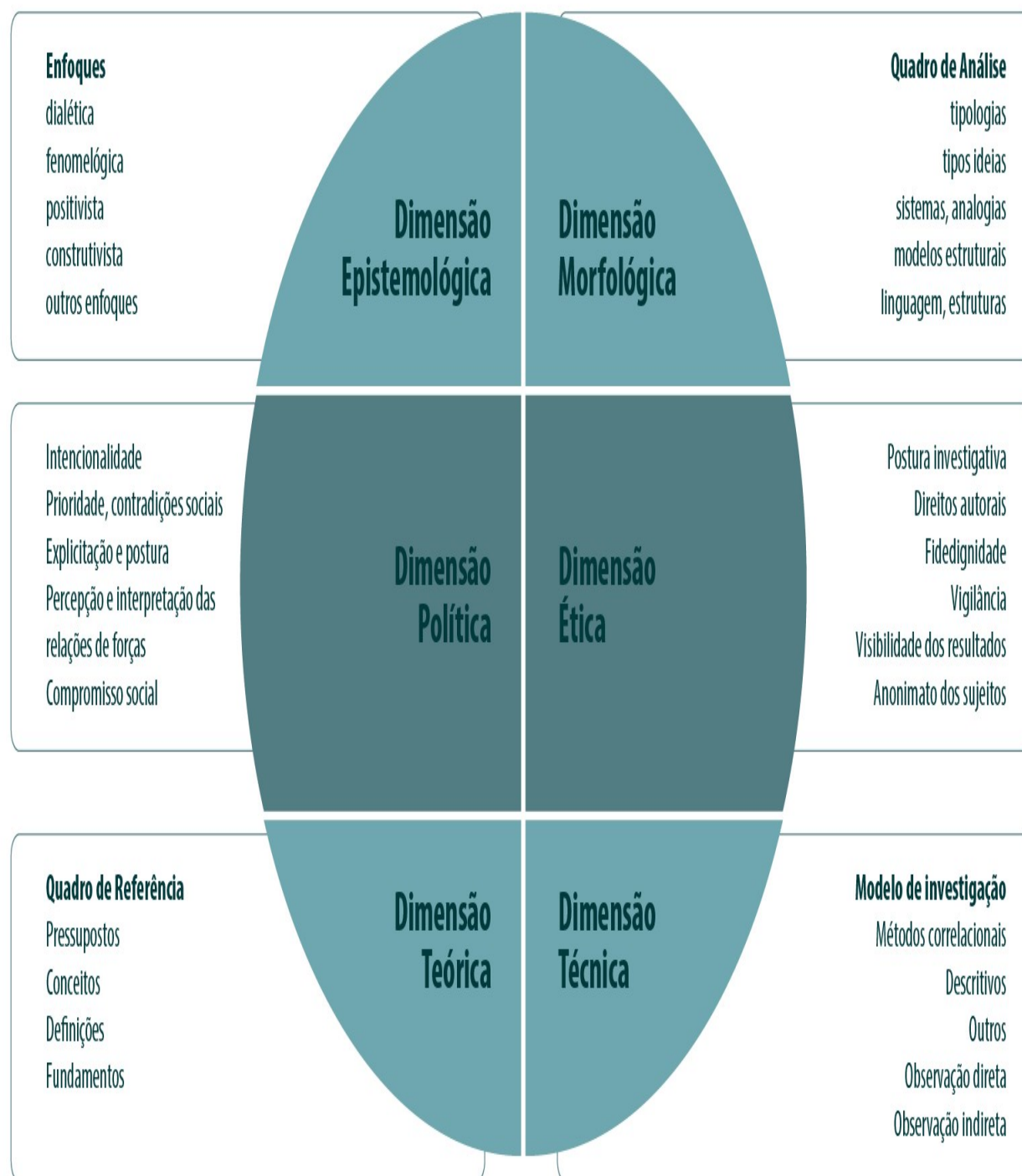
Assim, é mister pensar em um processo de ciência configurado em dimensões. As práticas científicas emergentes utilizam os dados científicos como insumos primários de investigações e podem apresentar resultados preliminares de pesquisa (BORGMAN, 2015). Parte assim do pressuposto que a estrutura é dinâmica e abrange diferentes dimensões (epistemológica, política, teórica, morfológica, técnica e teórica).

Por fim, entende-se que o movimento pela ciência aberta, em seu formato atual, reflete, na verdade, novos modos de pensar e de exercer a cientificidade do quarto paradigma, com a geração de novos modelos, técnicas, tecnologias que alterem o padrão e práticas dos paradigmas anteriores e proporciona o surgimento de novas hipóteses e inovação científica.

Percurso metodológico

O percurso metodológico adotado na pesquisa foi norteado pelas dimensões propostas por Bufrem (2015) que tem a base epistemológica do método quadripolar de Bruyne; Herman; Schoutheete (1977), conforme demonstrado na Figura 2.

Figura 2: Dimensões da Pesquisa



Fonte: Bufrem (2015).

A pesquisa é puramente qualitativa e adotou a análise de conteúdo de Bardin e o *Software NVivo* para apoiar na categorização de conteúdo. Baseado nas dimensões de pesquisa preconizadas por Bufrem, a análise partiu dos pressupostos e foi guiada por conceitos-chaves: ciência aberta, e-Science, quarto paradigma,

dados científicos, propriedade intelectual, ciência do *commons*, autoralidade colaborativa e recompensa autoral.

Quadro 1 – Categorias temáticas da análise de conteúdo

CATEGORIAS DE ANÁLISE	SUBCATEGORIAS	DIMENSÕES
Estruturação e concepções da Ciência Aberta e e-Science	Ciência Aberta e e-Science	Epistemológica e Teórica
Entendimento dos dados científicos sob a perspectiva do <i>commons</i>	Instrumentos Internacionais Regulatórios	Política e Legal
	Bem Intelectual Público e Comum	
	Domínio Público	
	Acesso Aberto	
Diretrizes político-normativas orientadas aos dados científicos	Princípios Normativo-legais para Dados Científicos:	Política e Legal
	Princípio Norteador 1: Proteção da Propriedade Intelectual	
	Princípio Norteador 2: Conformidade Legal	
	Princípio Norteador 3: Responsabilidade Formal	
	Diretrizes Político-legais brasileiras	
Sistematização e gerenciamento dos dados científicos	Modelo Ciclo de Vida dos dados	Morfológica e Técnica
Atribuição e Citação como Recompensa Autoral	Autoralidade na e-Science	Ética e Legal
	Autoria Colaborativa	
	Recompensa Autoral: atribuição, citação e responsabilização.	

Fonte: OLIVEIRA (2016).

A análise adotada foi a temática, construída por uma grade de categorias assentadas no *corpus* de análise, direcionada ao objeto de estudo e guiada pelos pressupostos e objetivos. O processo de codificação também denominado por unidade de registro, orientou na codificação e categorização dos conteúdos. Bardin (2011) afirma que a categorização é uma operação de classificação de elementos constitutivos de um conjunto por diferenciação e, em seguida, por reagrupamento com os critérios previamente definitivos. A análise temática compôs a sistemática do *corpus* de análise realizada no período de seis meses. Com base na categorização temática, as unidades de registro foram codificadas por temas, frases, parágrafos e sites.

O *corpus* de análise foi guiado pelo referencial teórico e pelas categorias temáticas: ciência aberta e propriedade intelectual. As fontes de pesquisa foram constituídas por artigos científicos (nacionais e internacionais), base de dados (*Elsevier*, *Proquest*, *Wiley Online Library*), legislações nacionais e documentos resultantes de diretrizes, normativas, iniciativas e instrumentos regulatórios (acordos, tratados e convenções).

Também fizeram parte do escopo portais governamentais, organizações não governamentais, repositórios de dados (científicos e governamentais), sites de instituições de ensino, de pesquisa e agências de fomento. Todo o levantamento bibliográfico foi por meio eletrônico, com predominância do conteúdo no idioma inglês, em virtude das fontes de pesquisa serem oriundas de diferentes países, em especial, os que já possuem práticas consolidadas de *Open Science*.

O escopo brasileiro foi analisado por meio de legislações específicas constituídas pela Lei de Direito Autoral (LDA), Lei de Acesso à Informação (LAI), decretos, instruções normativas e algumas iniciativas no âmbito dos órgãos governamentais e instituições científicas voltadas para o acesso aberto, dados abertos (governamentais e científicos). As conexões temáticas resultaram em categorias e subcategorias dispostas no Quadro 1, direcionadas à constituição das dimensões da Ciência Aberta. Para efeito deste artigo não foi contemplado a categoria Propriedade Intelectual.

Para a execução dos procedimentos de codificação e categorização foi utilizado o *Software NVivo* (versão 11 Mac). Os resultados obtidos a partir dos procedimentos gerados no *Software NVivo* ajudaram a constituir a análise de conteúdo e gerar representações de categorias e subcategorias(*nodes*).

No quadro acima, as categorias de análise foram selecionadas para entender as 6 (seis) dimensões propostas para pesquisas estabelecidas na essência da Ciência Aberta e que resultaram no modelo conceitual proposto no estudo.

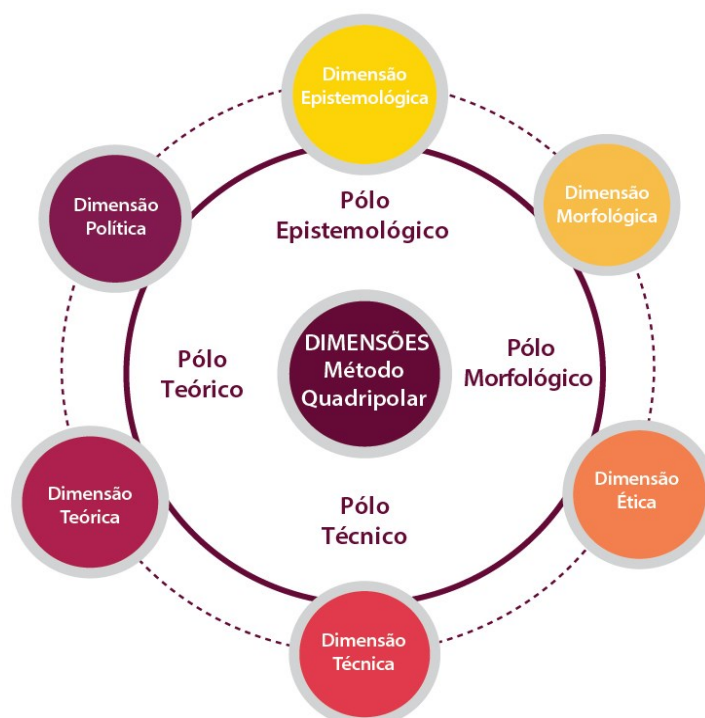
Descrevendo as dimensões da pesquisa

A literatura científica internacional aponta para as vantagens da pesquisa colaborativa aberta e muitas iniciativas e experiências estão em andamento. Nessa direção as dimensões de pesquisa que envolvem a Ciência Aberta foram pensadas como uma estruturação conceitual para que pesquisadores, universidades, agências de fomento e instituições de pesquisa observem os aspectos epistemológicos, políticos, teóricos, morfológicos, éticos e legais sejam abordados nas pesquisas científicas nos moldes da pesquisa aberta e colaborativa.

O modelo conceitual tem como pressupostos, o Método quadripolar (Bruyne; Herman; Schoutheete, 1977) que adota 04 (quatro) polos de pesquisa e as Dimensões de pesquisa (Bufrem, 2015) que nortearam a estruturação multidimensional em 6 (seis) dimensões, conforme demonstrado na Figura 3.



Figura 3: Dimensões da Ciência Aberta



Fonte: OLIVEIRA et al. (2019).

Dimensão epistemológica

Essa dimensão delimita o *constructo*, o objeto de estudo e a problemática de investigação. Os pesquisadores exprimem e exercem uma vigilância crítica e a objetivação (produção e explicitação do objeto e problemática) da pesquisa (BRUYNE; HERMAN; SCHOUTHEETE, 1977). A dimensão discursiva não remete diretamente ao conteúdo propriamente dito, mas à apreensão de ideias, linguagens e reflexões sobre o objeto de estudo.

A dimensão epistemológica exerce o domínio sobre o objeto científico em permanente construção e reformulação. A convicção e atitude científica do pesquisador sobre o objeto de estudo conduz a sua aceitação ou crítica com a produção de novos sentidos e significados sobre a ciência contemporânea.

É nessa dimensão que se expressa as reflexões: dos paradigmas da ciência observacional, experimental e computacional e do quarto paradigma da *e-Science*; os dados científicos como bens intelectuais; e, legislações e iniciativas políticas e normativas frente ao objeto de estudo.

Dimensão política

A ênfase dada a essa dimensão está associada ao domínio específico que abrange a Ciência. A noção do termo política “pertence ao domínio do conhecimento prático, e é de natureza normativa, estabelecendo os critérios da justiça [...]”

(JAPIASSÚ; MARCONDES, 2006, p. 220).

O arcabouço legal é insuficiente no provimento de regulamentações e instrumentos normativos que abranja as necessidades advindas da sociedade contemporânea, digital e compartilhada. Na ciência moderna os instrumentos jurídicos e as legislações específicas utilizam-se dos instrumentos políticos para dar sustentabilidade às estratégias e diretrizes desenvolvidas e adotadas para a promoção da pesquisa científica de forma acessível, aberta, gratuita e sustentável.

A estrutura política utiliza-se de instrumentos normativos e regulatórios para dar sustentabilidade às diretrizes político-normativas. Tais instrumentos são princípios, prescrições e regras que conduzem às decisões em um determinado contexto, os quais exercem um poder simbólico e normativo. A ‘norma’ serve como um “conjunto de prescrições relativas, as quais podemos decidir se algo é verdadeiro ou falso [...]” (JAPIASSÚ; MARCONDES, 2006, p. 220), ajudando a nortear os direcionamentos em um determinado domínio específico.

A dimensão política visa direcionar as diretrizes, normativas e políticas para a execução de pesquisas científicas abertas.

Dimensão teórica

A dimensão teórica orienta o pesquisador na elaboração de conceitos, pressupostos ou hipóteses, bem como na interpretação de fatos ou fenômenos e define provisoriamente soluções de pesquisa (BUFREM, 2013). A pesquisa progride concomitante ao processo de elaboração do referencial teórico. As noções preliminares que partem do senso comum e se estabelecem sistematicamente no escopo teórico são baseadas em conceitos, definições, enunciados e regras metodológicas direcionadas pela dimensão teórica.

Nesse estudo, a dimensão é direcionada para explicar e compreender o *constructo* teórico do objeto de estudo, pressupostos e conceitos norteadores em toda a pesquisa. As indagações postas no problema e objetivos são investigadas, refletidas e argumentadas na construção teórica.

A teoria é constituída por um ‘feixe de conceitos’ ou sistema conceitual que reflete um conjunto coerente de conceitos (MACHADO, 2007). É uma visão sistemática e encapsulada de conceitos que forma um todo coerente. O conceito, dessa feita, é a:

existência de uma questão, a formulação de um problema, a teoria apresenta determinada resposta, sugere uma solução. Privilegiar o conceito significa valorizar a ciência como processo. Esse aspecto dinâmico que caracteriza o conceito - e faz da ciência o domínio operatório [...] (MACHADO, 2007, p. 21).

O conceito é, pois, a abstração mental que se propõe em diferentes abordagens à reconstrução e à ressignificação de um contexto específico. Os conceitos devem ser formulados por nós – não os recebemos prontos – possibilitam o trabalho científico – a mediação do saber.

Dimensão morfológica

A dimensão morfológica é a estruturação e objetivação do objeto científico, o qual, adquire um espaço de representação. A representação formal do objeto científico, compreende: exposição do objeto do conhecimento, estilo pelo qual o investigador expressa os resultados e construção dos modelos de interpretação (BUFREM, 2013; SILVA, 2016).

Os resultados da investigação são formalizados e representam o objeto científico estudado e explica todo o processo de pesquisa e análise científica ao redor do objeto. É o pensar do objeto num espaço configurativo, um espaço qualquer de representação, por articular os conceitos, os elementos, as variáveis numa arquitetônica mais ou menos rigorosamente construída (BRUYNE et al., 1977). O objeto passa a ser configurado em uma articulação teórica, operacional, representacional, estrutural e prática.

A dimensão morfológica é composta pelos modelos científicos que estão presentes nas práticas da *e-Science* e reflete também a representação do modelo conceitual. No referencial teórico é possível analisar a dimensão morfológica sob o prisma do modelo de ciclo de vida dos dados e infraestrutura tecnológica. Tal modelo apresenta uma sistemática de estágios que estrutura as práticas no gerenciamento de dados científicos no ambiente da *e-Science*.

Por fim, a dimensão morfológica se constitui nas soluções técnicas e tecnológicas, articulando teorias, modelos e inferências para representar o objeto de estudo.

Dimensão ética

A dimensão ética é refletida pela preocupação com a fidedignidade, veracidade e honestidade que o pesquisador deve proceder durante o processo de pesquisa e os resultados finais. Nas práticas de pesquisa aberta, a preocupação é intensificada pela conduta ética, comportamento, honestidade, fidedignidade e veracidade científicas de pesquisadores, colaboradores e instituições. A postura e conduta ética investigativa são aspectos importantes que devem ser vigiados no processo de compartilhamento e colaboração científica.

Nesta pesquisa dois espectros foram pensados: ético-legal e recompensa autoral. O primeiro reflete as questões legais dos dados científicos de natureza pessoal, sigilosa e restrita, que na maioria das vezes obedecem instrumentos normativos ou legislações específicas. Paira nessa reflexão os critérios de veracidade, fidedignidade, autenticidade, sigilo e anonimização que estão presentes no escopo ético-legal.

O segundo espectro aponta para a recompensa autoral. A mudança no *status* da autoralidade dentro da *e-Science* intensifica preocupações legais quanto as condições e restrições de acesso, uso e reuso associadas à garantia de atribuição ao autor em relação aos dados primários da pesquisa.

Isto posto, a dimensão ética contemplada durante o desenvolvimento da pesquisa, a conduta baseada na fidedignidade, veracidade e honestidade para a feita do conteúdo, além de direcionar às reflexões ético-legais nas práticas

científicas emergentes.

Dimensão técnica

A dimensão técnica estabelece a relação entre a construção do objeto científico e a realidade através das abordagens, sistemáticas e operações técnicas. Os dados sobre o mundo real são obtidos mediante estratégias de pesquisa, técnicas de coleta de dados e evidências pertinentes à problemática da investigação, aos objetivos da pesquisa, ao universo e a amostra (BUFREM, 2013; SILVA, 2016).

Possui a finalidade de explicar procedimentos, técnicas, experimentações que são adotadas no decorrer da pesquisa. É quando o objeto de estudo entra em contato com a realidade objetivada e se materializa por meio de instrumentos, dispositivos metodológicos, teorias formuladas, conceitos operacionais e modelos representacionais.

Conclusão

A pesquisa concluiu que a Ciência Aberta é uma abordagem para investigações no contexto da comunicação científica aberta e da ciência-cidadã. É um termo derivado da expressão *Open Science* que surgiu no século XVI após a revolução pós-renascentista, com a apropriação dos bens intelectuais e ressurgiu nas últimas décadas como um grande esquema emergido do movimento de acesso aberto.

A Ciência Aberta se constitui como um termo guarda-chuva, amplo e multidimensional, que inclui acesso livre às publicações científicas, dados científicos abertos, documentos de pesquisa e ferramentas científicas abertas. É uma ciência voltada para grandes conjuntos de dados que são gerados em sofisticadas bases de dados e *software* que agrega uma ampla gama de simulações, metodologias e modelos efetivados em diversas áreas do conhecimento.

Os modelos científicos apresentam possibilidades que explicitam mais claramente a pesquisa desenvolvida, seus métodos, seu material, os dados e conhecimentos gerados, além de revelar aspectos intrínsecos da pesquisa científica, a exemplo da intenção do pesquisador e a avaliação dos resultados de forma mais eficaz e interativa.

Sob o ponto de vista dos dados científicos, apoia o compartilhamento e colaboração para uso, reuso e reprodutibilidade. O reuso possibilita dar continuidade a uma determinada pesquisa, gerar novas hipóteses, inferências, direcionamentos e alcançar novas descobertas a partir dos dados primários de pesquisa de outros pesquisadores. Esse processo de reuso é possível quando ocorre a efetiva colaboração científica.

Assim, a Ciência Aberta pensada como um guarda-chuva multidimensional abrange políticas, diretrizes e normas que envolvem novas práticas metodológicas, técnicas e tecnológicas que dão sustentabilidade às dinâmicas científicas da ciência contemporânea.

Referências

ABDO, A. H. Direções para uma academia contemporânea e aberta. In: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Org.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015. Disponível em: <https://www.cienciaaberta.net/ciencia-aberta-questoes-abertas-o-livro/>. Acesso em: 19 set. 2016.

ALBAGLI, S. Ciência aberta em questão. In: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Org.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015. Disponível em: <https://www.cienciaaberta.net/ciencia-aberta-questoes-abertas-o-livro/>. Acesso em: 19 set. 2016.

APPEL, André Luiz. **A e-Science e as atuais práticas de pesquisa científica**. Orientadora: Profa. Dra. Maria Lucia Maciel; Coorientadora: Profa. Dra. Sarita Albagli. Rio de Janeiro, 2014. 88 f. Dissertação (Mestrado em Ciência da Informação) – Escola de Comunicação, Universidade Federal do Rio de Janeiro; Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2014.

AUSTRALIAN NATIONAL DATA SERVICE. **Citation and identifiers**. In: H2020 Model Grant Agreements: H2020 General MGA, Australian, Oct. 2015. Disponível em: <http://www.andis.org.au/working-with-data/citation-and-identifiers>. Acesso em: 29 set. 2016.

AVIAN KNOWLEDGE NETWORK. **Data from: A Quantitative Survey of local adaptation and fitness trade-offs**. Disponível em: <http://datadryad.org/resource/doi:10.5061/dryad.315>. Acesso em: 16 jun. 2015.

BARACAT, A. C.; RIGOLIN, C. C. D. Proteção do Conhecimento e Movimento Open Access: Discussões no Âmbito da Organização Mundial da Propriedade Intelectual. **Liinc em Revista**, Rio de Janeiro, v. 8, n. 2, p. 348-364, set. 2012. Disponível em: <http://liinc.revista.ibict.br/index.php/liinc/article/viewFile/487/376>. Acesso em: 20 maio 2016.

BELL, G. Foreword. In: HEY, A. J. G.; TANSLEY, Stewart; TOLLE, K. M. **The Fourth Paradigm: Data-intensive Scientific Discovery**. Redmond, Wash: Microsoft Research, University of North Texas 2009. Disponível em: <http://digital.library.unt.edu/ark:/67531/metadc31516/>. Acesso em: 27 set. 2016.

BORGMAN, C. L. **Big Data, Little Data, No Data: Scholarship in the Networked World**. Cambridge, MA: MIT Press, 2015. Disponível em: <https://mitpress.mit.edu/big-data-little-data-no-data>. Acesso em: 16 dez. 2015.

BORGMAN, C. L. Data, Scholarship, and the eHumanities. **eMagazine**, v. 4, 2014. Disponível em: <http://ehumanities.leapress.com/emagazine-4/openingsscherm4/>. Acesso em: 17 abr. 2014.

_____. **From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World**. London: MIT Press, 2003. Disponível em: <https://mitpress.mit.edu/books/gutenberg-global-information-infrastructure>. Acesso em: 19 abr. 2016.

_____. **Research data: who will share what, with whom, when, and why?** In: CHINA-NORTH AMERICAN LIBRARY CONFERENCE, 5., Beijing, Sep. 2010. Disponível em: <https://works.bepress.com/borgman/238/>. Acesso em: 20 maio 2015.

_____. **Scholarship in the digital age: information, infrastructure, and the Internet.** Cambridge: The MIT Press, 2007.

BRUYNE, P.; HERMAN, J.; DE SCHOUTHEETE, M. **Dinâmica da pesquisa em ciências sociais: os pólos da prática metodológica.** Tradução de Ruth Joffily. Rio de Janeiro: Francisco Alves, 1977.

BUFREM, L. S. Configurações da pesquisa em ciência da informação, **DataGramaZero: Revista de Informação**, v. 14, n. 6, p. 1-15, dez. 2013.

CARBONI, G. C. Aspectos da teoria da função social do direito de autor. In: TIMM, L. B.; MACHADO, R. B. (Org.). **Função Social do Direito.** São Paulo: Quartier Latin, 2009.

_____. Direito autoral e acesso ao conhecimento: em busca do equilíbrio. **Revista Juris**, Faculdade de Direito da FAAP, São Paulo, v. 1, p. 32-98, jan./jun. 2009.

COMITÊ GESTOR DA INTERNET NO BRASIL. **Dados Abertos Governamentais.** Tradução de: opendatamanual.org. São Paulo: CGI.br, 2011. Disponível em: http://www.w3c.br/pub/Materiais/PublicacoesW3C/Manual_Dados_Abertos_WEB.pdf. Acesso em: 20 maio 2016.

_____. **Manual dos dados abertos: governo.** Tradução de: opendatamanual.org. São Paulo: CGI.br, 2011. Disponível em: <http://www.w3c.br/pub/Materiais/PublicacoesW3C/dados-abertos-governamentais.pdf>. Acesso em: 20 maio 2016.

EAVES, David. **The Three Laws of Open Government Data.** Sep. 11, 2009. Disponível em: <https://eaves.ca/2009/09/30/three-law-of-open-government-data/>. Acesso em: 25 maio 2015.

EUROPEAN COMMISSION. **Guidelines on the Implementation of Open Access to Scientific Publications and Research Data in Projects supported by the European Research Council under Horizon 2020**, Version 3.1, 25 August 2016. Disponível em: https://erc.europa.eu/sites/default/files/document/file/ERC_Guidelines_Implementation_Open_Access.pdf. Acesso em: 19 set. 2016.

_____. **H2020 Programme: Multi-Beneficiary General Model Grant Agreement, Version 3.0**, 20 Jul. 2016. Disponível em: https://erc.europa.eu/sites/default/files/document/file/ERC_Guidelines_Implementation_Open_Access.pdf. Acesso em: 19 set. 2016.

GRAY, J. eScience: a transformed scientific method. In: HEY, T.; TANSLEY, S.; TOLLE, K. (Ed.). **The Fourth Paradigm: Data-Intensive Scientific Discovery.** Redmond, Washington. UNT Digital Library. Palestra de Jim Gray em 11 de janeiro de 2007, cedida a NRC-CSTB 1, em Mountain View, CA. Disponível em: <http://digital.library.unt.edu/ark:/67531/metadc31516/>. Acesso em: 25 abr. 2016.

KOWALCZYK, S.; SHANKAR, K. Data Sharing in the Sciences. **Annual Review of Information Science and Technology**, Bloomington, Indiana University, Chapter 6, p. 247-294, 2014. Disponível em: https://kalpanashankar.files.wordpress.com/2010/06/arist_data_sharing.pdf. Acesso em: 9 dez. 2015.

MARQUES, F. Ciência no palheiro: Programa em eScience busca extrair novos conhecimentos em meio a volumes gigantescos de dados. **Revista Pesquisa FAPESP**, São Paulo, ed. 215, jan.



2014. Disponível em: <http://revistapesquisa.fapesp.br/2014/01/13/ciencia-palheiro/>. Acesso em: 19 mar. 2014.

MEDEIROS, J. S.; CAREGNATO, S. E. Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica. **Liinc em Revista**, Rio de Janeiro, v. 8, n. 2, p. 311-322, set. 2012. Disponível em: <http://liinc.revista.ibict.br/index.php/liinc>. Acesso em: 17 jan. 2014.

OLIVEIRA, A. C. S. de; SILVA, E. M. Ciência Aberta: dimensões para um novo fazer científico. **Informação & Informação**, Londrina, 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/issue/view/1143>. Acesso em: 20 set. 2016.

OLIVEIRA, Adriana Silva de Oliveira. **Desvendando a autorialidade colaborativa na e-science sob a ótica dos direitos de propriedade intelectual**. 2016. 300 f. Dissertação (Doutorado em Ciência da Informação) - Universidade Federal da Paraíba, João Pessoa, 2016. Disponível em: <https://repositorio.ufpb.br/jspui/handle/tede/8849>. Acesso em: 13 maio 2019.

TENOPIR, C. et al. Data Sharing by Scientists: Practices and Perceptions. **Plos One**, v. 6, n. 6, p. 1-21, 29 jun. 2011. Disponível em: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101>. Acesso em: 25 ago. 2016.

Ciencias abierta y ciudadana: por la visibilidad de las mujeres en la investigación

Hevellin Estrela¹, Suely Henrique de Aquino Gomes², Geisa Müller de Campos Ribeiro³, Mayllon Lyggon Oliveira⁴

¹ Faculdade de Informação e Comunicação. Universidade Federal de Goiás (FIC-UFG). Email: hevellinestrela@gmail.com

² Faculdade de Informação e Comunicação. Universidade Federal de Goiás (FIC-UFG). Email: suelyhenriquegomes@gmail.com

³ Faculdade de Informação e Comunicação. Universidade Federal de Goiás (FIC-UFG). Email: geisamuller@hotmail.com

⁴ Faculdade de Informação e Comunicação. Universidade Federal de Goiás (FIC-UFG). Email: mayllon.lyggon@gmail.com

Resumo

A partir da compreensão de que uma ciência aberta e cidadã deve ser também uma ciência inclusiva, e não somente voltada para a ampla disponibilização de dados e resultados de pesquisa, e que é cada vez mais crescente os estudos que apontam para a invisibilidade das mulheres na ciência, em decorrência, principalmente, de uma alegada falta de produtividade das pesquisadoras quando comparada com seus colegas do sexo oposto, este trabalho tem como objetivo verificar essa alegação a partir da análise das produções acadêmicas científicas de professoras e de professores vinculados a cursos de pós-graduação, nível doutorado, em Ciências Agrárias do Centro-Oeste do Brasil. A produção analisada compreende o período de 2013 a 2017. A apresentação desses dados reforça a investigação das relações de gênero e ciências nas atividades acadêmicas científicas, tratando da inserção da mulher na ciência e do mercado de trabalho no ensino e na pesquisa. A metodologia utilizada é a qualitativa. Para aferir diferenças estatisticamente significantes entre a produção acadêmica entre professoras e professores da população estudada usou-se o teste estatístico “t de *student*” e *chi-quadrado*. Os dados foram extraídos da plataforma Lattes, utilizando a técnica de *webscraping*. A partir da obtenção dos dados foi possível realizar uma análise inferencial. Os resultados mostram que as mulheres são minoria nessa área, mas que elas são responsáveis por 51% da produção das atividades acadêmicas científicas desenvolvidas na área estudada. Demonstra ainda que as atividades as quais as mulheres estão mais atreladas são àquelas que estão relacionadas ao educar e direcionar, como as orientações de trabalhos. Além disso, constatou-se que a partir dos 50 anos, aumenta consideravelmente a média de produção intelectual das mulheres (44 artigos por docente) e, conseqüentemente, o percentual de detentoras de bolsa produtividade (60% das mulheres com idade igual ou superior a 60 anos são bolsistas, enquanto que somente 18 % dos homens nessa faixa etária recebem bolsas produtividade), mas essa tendência

não é estatisticamente significativa. Os resultados aqui apresentados a necessidade de ciência aberta também inclusiva e não meramente um movimento de promoção de acesso a dados e produções científicas.

Palavras-chave: Gênero e Ciência. Ciências Agrárias. Ciência aberta. Mulheres na ciência.

Introdução

O movimento da ciência aberta vem em contraposição à lógica da produção científica industrial. Trata-se de pôr freio à “privatização do saber, que é central na atual relação entre ciência, tecnologia e mercado”, sendo a década de 1980 emblemática na expansão dessa lógica de uma ciência capitalista (CASTELFRANCHI, 2008, p. 94).

Esse modo de produção do conhecimento, denominada por Etzkowitz (2002) de tríplice hélice, está imbuído “de promover maior acumulação de capital para grupos já economicamente privilegiados” (GOMES; GOMES, pré-print). Em prol da competitividade - garantida pelo sigilo-segredo industrial, instaura-se a tendência à restrição de acesso ao conhecimento que foi, em muitos casos, produzido a partir de financiamentos públicos e, no caso do Brasil, nas dependências de uma instituição pública.

A ciência aberta vem na contramão dessa tendência da privatização dos resultados científicos e procura estabelecer uma cultura de produção coletiva do conhecimento e de amplo compartilhamento e publicização de dados científicos, informações, resultados de pesquisa, sem as barreiras que limitem seu acesso e reutilização (ALBAGLI; CLINIO; RAYCHTOCK, 2014). É um movimento que surge em prol de uma ciência cidadã e democrática que preza pelo envolvimento do público leigo na definição de prioridades de pesquisa e no estabelecimento dos rumos da ciência.

No entanto, para envolver o público leigo no empreendimento científico é necessário educar para a ciência e “romper o desinteresse e desilusão, principalmente dos jovens, com a coisa pública, com os políticos e, porque não dizer, com a própria ciência” (GOMES; GOMES, pré-print). Nesse sentido, a discussão de consolidação de um projeto de ciência cidadã obriga-nos a trazer o debate para o campo educacional e político, mesmo porque, como nos propõem Gomes e Gomes (pré-print): “isolados, eles não serão suficientes para romper com a lógica capitalista e de inevitabilidade instaurada no campo científico-tecnológico”.

Sendo a ciência fruto de uma determinada formação histórico-social, não tem como a mesma ser apolítica: a ciência é uma atividade politizada, queiram ou reconheçam seus praticantes ou não, e essa constatação coloca em xeque os princípios de neutralidade e universalidade que alicerçam e são tão caros à constituição da ciência moderna.

Araújo e Facchini (2018, p. 1) nos convidam a “refletir sobre a efetividade dos avanços nas condições de vida, na autonomia e liberdade das mulheres, bem como na superação das situações de violência e desigualdade nas quais se inseriram”. E aqui

incluem-se as interrogações sobre as condições de desigualdades instauradas no interior da própria ciência.

Neste contexto, as discussões sobre a temática mulher na ciência têm crescido ao longo dos anos partir de diversas abordagens e campos de conhecimento, como aquelas promovidos por Silva (1998), Leta (2003), Olinto (2011), Cruz (2011), dentre outras. Entretanto, ainda existem alguns questionamentos sobre a participação da mulher no fazer científico. As evidências sugerem que há alguns aspectos relacionados à tradição cultural do fazer científico que sempre esteve associado ao masculino, como exemplo, alguns princípios norteadores da Ciência Moderna tais como neutralidade e objetividade.

Essas discussões expressam que a inserção da mulher na ciência é um problema de relações de gênero, consequência de um processo histórico e cultural que estabelecem lugares que são valorados de forma hierárquica para mulheres e para homens. Mesmo com pouca visibilidade, sabe-se que mulheres contribuíram para o crescimento e evolução do campo científico durante a história da consolidação da pesquisa. São muitos deslocamentos teóricos produzidos por muitas mulheres. No entanto, essas discussões ocorrem em paralelo e independentes àquelas sobre a constituição de uma ciência aberta e cidadã.

Nesse sentido, a presente pesquisa vem nessa dupla perspectiva, talvez pretenciosa: a) inserir no debate sobre ciência aberta e cidadã a politização da questão gênero e de outras minorias na ciência na tentativa de ampliar o seu conceito, uma vez que uma ciência que se pretende cidadã não deve fugir desse debate. Essa intenção é aqui colocada como pressuposto para o nosso segundo intento; b) trazer os resultados de análise comparativa entre as produções acadêmicas científicas das professoras e dos professores vinculados aos cursos de Pós-Graduação da região Centro-Oeste vinculados a área de Ciências Agrárias para constar as desigualdades de gênero nesse campo científico, com o intuito maior de verificar as desigualdades de gênero ali instaladas.

A justificativa para escolha da área de Ciências Agrárias dá-se devido a este ser um dos primeiros cursos do ensino superior no Brasil (CAPDEVILLE, 1991), além de ser a área de maior produção científica e que recebeu o maior investimento para essa produção conforme os dados do CNPq (2017) e a escolha pela região dá-se pelo fato que o setor agropecuário foi o que teve maior participação no produto interno bruto – PIB entre os anos de 2014 a 2017, representando 14,8% do PIB. E em Goiás, segundo dados da Secretaria de Estado de Gestão e Planejamento - SEGPLAN (2015, p. 9), este é o setor que mais se destaca no estado e contribui com 7,2% da agropecuária nacional.

Ciência aberta e cidadã: tensionando conceitos

As novas dinâmicas de produção e circulação do conhecimento, da informação e da cultura, assim como os novos processos de participação científica apresentam a temática ciência aberta tanto para o universo acadêmico, como para atores sociais que se preocupam com a democratização do conhecimento e da informação. (ALBAGLI; MACIEL; ABDO, 2015).

Pode-se dizer que a ciência aberta é uma passagem do movimento de acesso aberto e outras frentes, como: dados científicos abertos, ferramentas científicas abertas,

hardware científico aberto, cadernos científicos abertos e wikipesquisa, ciência cidadã e educação aberta. Se constitui como “um termo guarda-chuva, que engloba diferentes significados, tipos de práticas e iniciativas, bem como envolve distintas perspectivas, pressupostos e implicações”. Estão incluídas desde a disponibilização gratuita dos resultados da pesquisa, que pode ser chamado de acesso aberto, até a valorização e a participação direta de não cientistas e não especialistas no fazer científico. (ALBAGLI; CLINIO; RAYCHTOCK, 2014, p. 435).

Acessar informação é, de fato, um direito do cidadão, conforme expressa a International Federação de Library Associations e Instituições - IFLA (2014) e nossa constituição. Contudo, algumas problemáticas permeiam essa questão envolvendo agências financiadoras de pesquisas, editoras, pesquisadores(as) e atores distintos.

A verdade é que o acesso a informações sem impeditivos de ordem financeira ou de licenças é uma demanda cada vez constante. O cidadão quer ter acesso, quer manipular, ver e interagir com os conteúdos. Nesse sentido, a internet desafia os métodos tradicionais de publicação utilizados pela comunidade científica sustentado em assinaturas e dependência das publicações científicas em relação a editores privados que acabam por inviabilizar o acesso de forma livre.

Estas discussões apresentam pontos de vistas distintos e em muitos casos conflituosos. Em primeiro, destaca-se a vertente de socialização do conhecimento, da informação e da cultura e seus mecanismos de privatização e captura. Em segundo, destaca-se a vertente da abrangência do próprio significado da ciência aberta que visa a interlocução da ciência com outros segmentos sociais e outros tipos de saberes, no amplo espectro de possibilidades e espaços de produção colaborativa, interativa e compartilhada do conhecimento (ALBAGLI; MACIEL; ABDO, 2015).

Os mesmos autores dizem que

A ciência aberta implica superar a perspectiva de pensar a ciência a partir da sua produtividade intrínseca. Implica o abalo de hierarquias, de fontes estabelecidas de autoridade e reputação, colocando foco nas relações entre ciência e poder, e, mais amplamente entre saber e poder (ALBAGLI; MACIEL; ABDO, 2015, p. 10).

Compreender a discussão de ciência aberta na perspectiva de processos de mudança social é de suma importância e instaura a necessidade do entendimento de uma ciência cidadã. Ou seja, é pensar a abertura da ciência em relações de saber poder e nas diferentes formas de produção de conhecimento.

Pode-se dizer que a ciência cidadã não é um conceito novo, mas suas discussões emergem a partir dos avanços das Tecnologias da Informação e comunicação, principalmente nos sentidos de produção e acesso da informação e conhecimento.

Parra (2015, p. 126) apresenta a ciência cidadã como,

[...] engajamento do público em geral nas atividades de pesquisa científica quando os cidadãos contribuem ativamente para a ciência, seja com seu esforço intelectual, com seu conhecimento local ou com suas ferramentas e recursos. Os participantes fornecem dados experimentais e instalações para pesquisadores, elaboram novas questões e criam conjuntamente

uma nova cultura científica. Enquanto agregam valor, os voluntários adquirem de forma instigante novos aprendizados e habilidades e um conhecimento mais profundo do trabalho científico.

Portanto, caracteriza-se como a tentativa de um outro paradigma de fazer ciência que aproxima o pesquisador da sociedade e prevê o engajamento da população na geração, resolução de problemas e análise de dados. Mais do que isso, apresentam novas formas de conhecer, instaura novas dinâmicas de produção de conhecimento, validação e difusão fazendo emergir formas de colaboração entre cientistas e pesquisadores não-acadêmicos, reinventando a dimensão política da ciência (PARRA, 2015).

Mas um dos grandes tensionamentos que tem permeado a reflexão da pesquisa, está em pensar como é possível apresentar o movimento de ciência aberta e cidadã, se o modo de produção científica, ou seja, a integração de um coletivo social no âmbito científico não discute a questão de gênero e de outras minorias na ciência? Como pensar o engajamento cidadão, reconhecer e dar visibilidade a outros sujeitos e práticas de conhecimento se no interior da própria ciência foi construído historicamente um modelo universal de “fazer ciência” associado ao masculino que invisibilizou “mulheres como sujeitos de conhecimento e do conhecimento” (SILVA, 2008)?

Para pensar a abertura da ciência em relações de saber poder e nas diferentes formas de produção de conhecimento deve-se abandonar a concepção de ciência como um campo despolitizado. A ciência como produto social traz para o seu interior uma estrutura de desigualdades na produção que exclui determinados grupos que dela fazem parte. Uma ciência cidadã e inclusiva deveria reconhecer e dar visibilidade às contribuições de grupos minoritários, seja considerando a raça, seja trazendo a questão de gênero à tona. Uma ciência cidadã e inclusiva tem que incluir a perspectiva dos “historicamente dominados”, conforme defende Silva (2008).

Devemos considerar e, quiçá, expandir o conceito de ciência aberta para torná-la também inclusiva, conforme preconiza Dagnino (2006) na proposição de seu modelo de estudos da ciência denominado pelo autor de CTS2. A CTS2 defende uma atuação direta daqueles descontentes que, de direito, querem que seus interesses e demandas sejam levados em consideração no processo de produção do conhecimento tecnocientífico e na definição de políticas em C&TI. A questão, portanto, não estaria no como o conhecimento é utilizado, mas no porquê, por quem e para quem o mesmo foi produzido; qual o projeto político ou regime social a que ele está subserviente, Agenda presente, tanto no movimento da Ciência Aberta, quanto nas críticas feministas à ciência, que trataremos brevemente a seguir.

As mulheres na ciência: a crítica feminista

Muitas mulheres contribuíram para grandes avanços e conquistas científicas, mas foram os movimentos feministas que proporcionaram novos sentidos às ações de mulheres e sua participação na vida social, política, econômica e cultural e científica. As lutas feministas provocaram rupturas e deslocamentos de questões pertinentes à moral,

à sexualidade, aos modos de feminilidade e à corporeidade e aos modos tradicionais de produção de subjetividade (RAGO, 2013).

No que tange o campo de produção do conhecimento científico, segundo Sadenberg (2002, p 1), “a crítica feminista tem avançado da mera denúncia da exclusão e invisibilidade das mulheres no mundo da ciência para o questionamento dos próprios pressupostos básicos da Ciência Moderna, virando-a de cabeça para baixo ao revelar que ela não é nem nunca foi ‘neutra’”. As críticas femininas à ciência colocam em xeque os pressupostos da neutralidade, universalidade e objetividade – valores caros para a ciência estruturada sob a batuta do patriarcalismo.

Schiebinger (2001, p. 37) reitera que a exclusão, minimização ou não reconhecimento dos feitos das mulheres na ciência “é um produto de centenas de anos”. O processo para trazê-las nesse contexto e dar visibilidade aos seus feitos exigem profundas mudanças na cultura, métodos e conteúdos da ciência. Os avanços feitos nessa direção ainda não devem ser motivos de júbilo nem prenúncio de que as mulheres terão êxito em um ambiente que inicialmente foi estruturado para excluí-las, nos alerta Schiebinger (2001). Segundo Silva (2008, p. 135) os preceitos que “balizaram a construção do conhecimento científico, [abstraíram], declaradamente, toda possibilidade de considerar as mulheres como sujeitos de conhecimento e do conhecimento”. Além disso, estruturas sociais, o desenvolvimento econômico do país, o sistema educativo e a ausência de sistemas que viabilizem a vida profissional e familiar da mulher, são elementos que podem dificultar o seu avanço nessa seara.

Bandeira (2008, p. 208) aponta que a premissa da crítica feminista à ciência parte do pressuposto histórico que a produção do conhecimento científico está ‘reservada’ aos homens, mas que isso não significa a exclusão das mulheres, porém explicita que as resistências às presenças de mulheres no campo científico são ainda instigantes, mesmo reconhecendo-se “que a existência de um sujeito universal já não é mais plausível, e isso vale tanto para o eu masculino como para a tardia individualidade feminina”. Nem sujeito, muito menos uma ciência universalizada.

A crítica feminista aos princípios da universalidade e, incluímos nesse bojo, o de uma ciência democrática é que, para assim ser considerada, deve-se incluir “a perspectiva dos dominados”, ou seja, não se deve ignorar ou evadir-se do debate sobre as questões de gênero, raça, etnicidade e classe nos modos de fazer ciência (SILVA, 1998, p.110). Extrapolamos essa posição para o movimento da ciência aberta e cidadã: para ser cidadã a ciência deve incluir a perspectiva dos dominados e travar o debate dessas questões na nova proposição desse modo de fazer ciência.

Dentre as grandes conquistas e avanços na ocupação de espaços no campo da ciência está o o acesso ao ensino. Nos dias atuais já é possível observar que as mulheres representam 70,6%, de matrículas na graduação, conforme os dados do Portal Brasil de 2016 e do censo da educação superior do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - INEP (INEP, 2017). As mulheres são também a maioria em cursos de pós-graduação (mestrado, mestrado profissional e doutorado), conforme os dados da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES apresentados na tabela 1:

Tabela 1: Ingressos, matriculados e concluintes em cursos de pós-graduação no Brasil – 2015.

Sexo	DOUTORADO		MESTRADO		MESTRADO PROFISSIONAL	
	MATRICULADO	TITULADO	MATRICULADO	TITULADO	MATRICULADO	TITULADO
FEMININO	54.491	10.141	66.439	26.443	13.529	4.376
MASCULINO	47.877	8.484	55.175	20.215	14.390	4.095
Total Geral	102.368	18.625	121.614	46.658	27.919	8.471

Fonte: CNPq (2015)

No entanto, há desequilíbrio da representatividade das mulheres nas diversas áreas do conhecimento: a maior parte concentra-se nas áreas de Humanas, e são pouco expressivas nas Exatas, o que sugere uma “inclusão com segregação” caracterizada pela “interdição ou desestímulo ao acesso feminino a certas áreas do conhecimento e profissões que se mantiveram como redutos masculinos” (BRITO; PAVANI; LIMA Jr., 2015, p 36).

Outro fenômeno relevante é diminuição da participação feminina em postos mais avançados das carreiras por elas abraçadas (AGRELLO e GRAG, 2009) fenômeno denominado de “Efeito Tesoura” porque corta a possibilidade das mulheres progredirem profissionalmente. Esse efeito não está confinado a determinado segmento do mundo do trabalho e pode ser observado, também, na academia, na ocupação de instâncias decisórias de Ciência e Tecnologia e na distribuição, por exemplo, de bolsas produtividade em Pesquisa, concedidas pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) aos profissionais reconhecida contribuições na produção do conhecimento (BRITO, PAVANI, LIMA Jr. 2015). No entanto, estudos

mostram quantitativamente que o número de artigos e fator H das mulheres é equivalente ou melhor que o dos homens, dependendo do nível (ARENZON et al., 2013). Ou seja, à medida que as mulheres avançam em suas carreiras há uma gradativa perda de espaço que não ocorre necessariamente por efeitos meritocráticos (BRITO, PAVANI, LIMA Jr. 2015, p 38).

Mas “para que aumentar a participação feminina em carreiras de C&T?” perguntam-se os autores acima citados (2015, p 38). Uma linha de argumentação desenvolvida por eles é que

povoar a ciência com mais mulheres não implica somente a produção de um corpo científico mais competente. Ampliar o debate sobre a participação de mulheres na ciência significa também pensar numa ciência diferente, inspirada e renovada por experiências de vida historicamente excluídas da produção científica e tecnológica (BRITO, PAVANI, LIMA Jr. 2015, p 39).

Metodologia

A pesquisa classifica-se como quantitativa e quanto aos objetivos como exploratória. O público alvo do estudo são os 162 docentes da região Centro-Oeste vinculados aos programas de pós-graduação (nível de doutorado) na área de Ciências Agrárias, ofertados pelas seguintes instituições: IFGOIANO, UEMS, UFG, UFGD, UFMT, UNB.

Os dados sobre as atividades acadêmico-científicas da população estudada foram extraídos da Plataforma Lattes¹ e a plataforma Sucupira², usando-se o *webscraping*. Para análise, após a quantificação dos dados fez-se a média por atividade acadêmica científica, considerando-se as variáveis sexo e faixa etária. Após a média, então aplicou-se o teste estatístico e análise inferencial dos dados extraídos.

Na extração dos dados, percebeu-se que havia vários outsiders, o que se fez necessário o uso da metodologia KDD (Knowledge Discovery in Databases ou Descoberta de Conhecimento nas Bases de Dados) que segundo Fayyad (1996) *apud* Camilo e Silva (2009, p. 4), que o define como "um processo não trivial de identificação de novos padrões válidos, úteis e compreensíveis".

Para a extração dos dados quantitativos dos docentes, utilizou-se a técnica *webscraping*, que tem se tornado comum no meio científico, e que literalmente quer dizer raspagem de dados da web. Marres e Weltevrede (2013) descrevem *webscraping* como uma técnica proeminente para coleta de dados online automatizada.

Escolheu-se a distribuição T-student e Chi-quadrado para procedermos a análise inferencial. Ferreira (2005, p. 10) o "objetivo desse tipo de análise "é obter uma afirmação acerca da população com base numa amostra". No caso da presente pesquisa, trabalhamos com o total da população (162 docentes pesquisadores).

Resultados e discussão

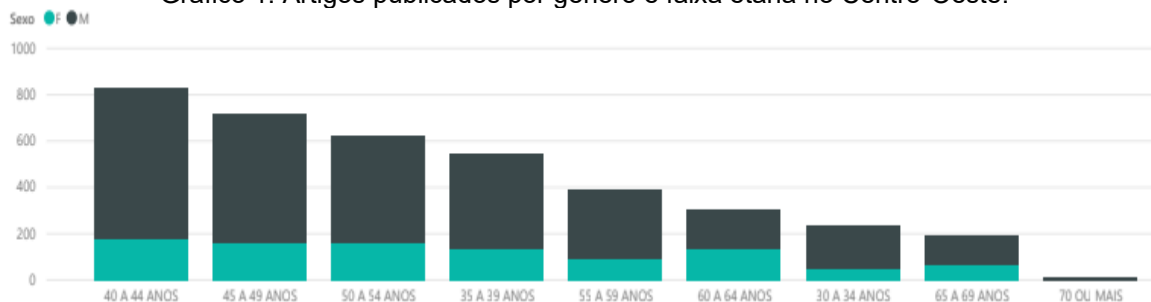
A maioria dos docentes nas pós-graduações em Ciências Agrárias na região Centro Oeste, segundo os dados levantados, é do gênero masculino, o que está de acordo com os dados do INEP (2017). As mulheres correspondem a 28% (45/162) dos docentes dos cursos analisados. No gênero feminino, considerando-se a faixa etária, a classe modal é de 45 a 49 anos (9 professoras) e no grupo masculino é de 40 a 44 anos (21 professores). Esses dados diferem daqueles apresentados no estudo do INEP que aponta 34 anos como idade predominante na docência superior brasileira.

¹ Plataforma virtual Lattes, mantida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ), é um sistema de informações que integra as bases de dados de currículos, grupos de pesquisa e instituições como registro da vida pregressa e atual de estudantes e pesquisadores do país. Informação disponível em: <http://lattes.cnpq.br/>

² A Plataforma Sucupira é uma base de referência do Sistema Nacional de Pós-Graduação (SNPG) utilizada para coletar informações, realizar análises e avaliações dos cursos de pós-graduação no país. Informação disponível em: <https://www.capes.gov.br/avaliacao/plataforma-sucupira>

Uma das atividades a ser desenvolvida pelos professores-pesquisadores é a publicação de artigos em revistas científicas. O gráfico 1 revela um maior número de publicação na faixa etária dos 40 aos 44 anos para ambos os sexos. No entanto, nota-se que diferença entre o número de artigos publicados da classe modal e as duas classes subsequentes é maior no gênero masculino que no feminino. A média de publicação dos homens 24,66 artigos publicados em revistas, enquanto a das mulheres é de 21,69.

Gráfico 1: Artigos publicados por gênero e faixa etária no Centro-Oeste.



Fonte: Os autores (2019)

Quando aplicamos o teste *t-student*, o mesmo aponta para diferença estatística entre a média de artigos completos publicados em revistas entre os gêneros ($<0,05$), ou seja, observa-se uma leve tendência dos homens publicarem, em média, mais artigos em revistas científicas que as mulheres. Porém, pelos dados expostos, o que se percebe é que “mesmo com todas as dificuldades impostas ao sexo feminino, conseguimos produzir cientificamente” (TENENTE, 2007, p. 73) o que, talvez nos falte, é o reconhecimento dessa produção científica.

O resultado do *t-student* para publicação de capítulos de livros é de 0,008 o que indica que há diferenças estatísticas entre as médias desse tipo de produção entre mulheres (1,47) e homens (1,97). Quando consideramos trabalhos completos publicados em anais de congressos, temos a média de 1,53 para as mulheres e de 4,16 para os homens. Constatamos diferença estatística também para essa atividade (*t-student* = 0,04)

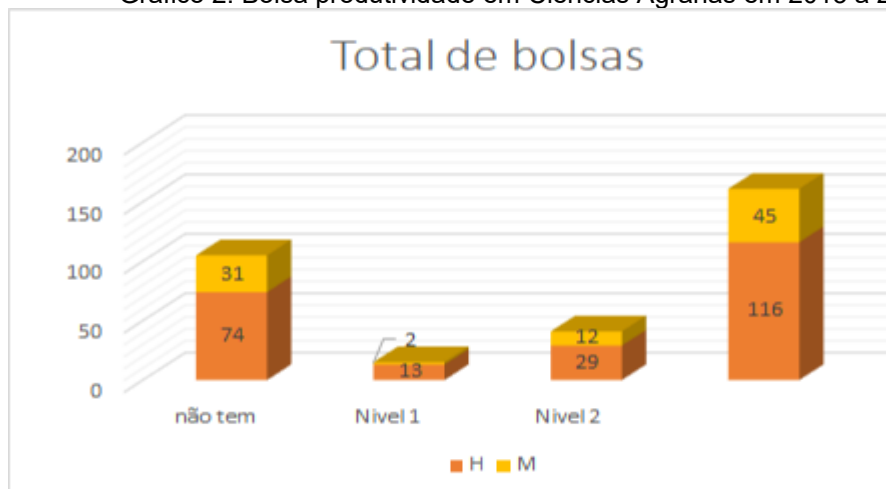
Sobre a publicação de resumo expandido em anais de congressos, a média para essa atividade é de 9,22 para as mulheres enquanto para os homens a média é de 11,47 e em relação a aplicação do teste *t-student* temos o resultado de 0,111, o que significa que não há diferenças estatisticamente significante entre homens e mulheres nesse item.

As atividades que as mulheres possuem a maior média são nas atividades consideradas como o ato de maternar que são as orientações e participações em bancas. Na participação de banca de Trabalhos de Conclusão de Curso -TCC a média das mulheres é de 29,71 e a média dos homens 26,30 e a aplicação do teste estatístico usado nessa pesquisa é de 0,05. E nas orientações e supervisões concluídas a média das mulheres é de 7,07 e a dos homens é de 6,06 (*t-student* é igual a 0,111).

A análise para identificar a distribuição das bolsas bolsa entre a população pesquisada mostra que do total de professores dos programas de pós-graduação (doutorado) em Ciências Agrárias do centro-oeste, 36,6% dos homens são bolsistas e 31,1% das mulheres estão nessa condição (gráfico 2). Quando procedemos análise por

faixa etária, isolando-se os grupos de homens e mulheres, constatamos que: do total de mulheres na faixa etária acima dos 60 anos, 60% têm bolsas (3-5)³, enquanto o percentual de homens bolsistas nesse mesmo intervalo de idade é de 18,5% (3-16); dos 50 a 59 anos: 38% das mulheres (5-13) e 36% dos homens (12;33) nesse intervalo são bolsistas; dos 40 a 49 anos: 20% das mulheres (3;15) e 40% dos homens (16;40) são bolsistas; dos 30-39 anos: 16,7% (2-12) das mulheres e 19,2% dos homens são bolsistas.

Gráfico 2: Bolsa produtividade em Ciências Agrárias em 2013 a 2017.



Fonte: Os autores (2019)

Os dados sugerem que à medida que as mulheres avançam na idade e permanecem na carreira (não se aposentam), elas podem se tornar mais competitivas. Uma provável explicação para esse fato é que a partir dos 50 anos, as mulheres se liberam dos cuidados com os filhos. Entretanto, não há diferenças estatisticamente comprovada entre a distribuição de bolsas por gênero, idade e categoria de bolsa ($\chi^2 = 0,67$) entre a população estudada.

Conclusões

Os resultados obtidos corroboram com a fala de Woolf (2018, p. 50) quando nos diz que as mulheres precisam mais do que educação, que é preciso ter “liberdade de experiência e que possam divergir dos homens sem receio de expressar suas diferenças”, pois se não mudarmos os resultados, ou no caso desse estudo, promover o acesso e a permanência na carreira científica sem considerar a jornada dupla das pesquisadoras. Sem políticas adequadas, o que se espera é a “perpetuação do domínio de um lado e, de outro, a servilidade” (WOOLF, 2018, p.51).

As diferenças estatísticas na produção científica entre homens e mulheres não nos exime de refletir sobre a provocação de Viegas, Gazire, Reis e Barbosa (2016, p. 12)

³ Apesar de parecer um percentual muito significativo quando comparados ao dos homens, a população analisada (5 pesquisadoras) é pequena para fazer aferições mais contundentes.

quando nos indaga o que é ser mulher na contemporaneidade? E nos respondem que “a categoria mulher, há pelo menos 30 anos não se encaixa na estrutura estanque de um subjetivo que as sujeita, mas sim em diferentes potências e deveres”.

Essas considerações repercutem os estudos de Leta (2003, p. 280) quando nos diz que: “as estatísticas mostradas anteriormente apontam que, se por um lado, as mulheres têm participado cada vez mais das atividades de C&T no Brasil, por outro, elas ainda não avançam em cargos e posições de destaque e reconhecimento” e essa situação é também percebida quando, após análise da composição do Comitê de Ciências Agrárias, constata-se a predominância masculina nessa instância de análise e decisória.

Os dados aqui apresentados que apontam para o que Brito, Pavini e Lima Jr. (2015, p. 36) denomina “uma inclusão progressiva caracterizada pela segregação, com interdição ou desestímulo ao acesso feminino a certas áreas do conhecimento e profissões que se mantiveram como redutos masculinos”. Que o movimento por uma Ciência aberta incorpore essa causa em seus princípios e agenda.

Referências

ALBAGLI, Sarita; CLINIO, Anne; RAYCHTOCK, Sabryna. Ciência aberta: correntes interpretativas e tipos de ação. **Liinc em Revista**, Rio de Janeiro, v. 10, n. 2, nov. 2014. Disponível em: <http://revista.ibict.br/liinc/article/view/3593/3072>

ALBAGLI, Sarita; MACIEL, Maria Lucia; ABDO, Alexandre Hannud. (Org). **Ciência aberta**: questões abertas. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015.

ARAUJO, Angela Maria, FACCHINI, Regina. Mulheres e direitos humanos no Brasil: avanços e desafios. **Jornal da Unicamp**. 2018. Disponível em: <https://www.unicamp.br/unicamp/ju/artigos/direitos-humanos/mulheres-e-direitos-humanos-no-brasil-avancos-e-desafios>. Acesso em: 7 dez. 2018.

BANDEIRA, Lourdes. A contribuição da crítica feminista à ciência. **Estudos Feministas**, Florianópolis, v. 16, n. 1, jan./abr. 2008. Disponível em: <http://www.scielo.br/pdf/ref/v16n1/a20v16n1>. Acesso em: 2 jan. 2019.

BARBOSA, M. C.; LIMA, B. S. Mulheres na Física do Brasil: Por que tão poucas? E por que tão devagar? In: YANNOULAS, S. C. (Coord). **Trabalhadoras**: análise da feminização das profissões e ocupações. Brasília: Editorial Abaré, 2013. Disponível em: http://www.oitcinterfor.org/sites/default/files/file_publicacion/trabajadoras.pdf. Acesso em: 13 jan. 2019.

BRITO, Carolina; PAVANI, Daniela Pavani; LIMA Jr., Paulo. Meninas na Ciência: atraindo jovens mulheres para carreiras de Ciência e Tecnologia. **Gênero**, v.16, n.1. p. 33 - 50, set./dez. 2015. Disponível em

<http://www.revistagenero.uff.br/index.php/revistagenero/article/view/744/411>. Acesso em 15 maio 2019.

CAMILO, Cássio Oliveira; SILVA, João Carlos da. **Mineração de dados: conceitos, tarefas, métodos e ferramentas**. 2009. Disponível em: http://www.portal.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-09.pdf. Acesso em: 15 maio 2019.

CASTELFRANCHI, Yuri. **As serpentes e o bastão: Tecnociência, neoliberalismo e inexorabilidade**. 2008. 392 f. Tese (Doutorado em Sociologia) – Instituto de Filosofia e Ciências Sociais da Universidade de Campinas, Campinas, 2008. Disponível em: <https://cteme.files.wordpress.com/2009/06/castelfranchijuri.pdf>. Acesso em: 2 mar. 2019.

CAPDEVILLE, Guy. O ensino superior agrícola no Brasil. **R. bras. Est. pedag.**, Brasília, v.72, n.172, p.229-261, set./dez. 1991. Disponível em: <http://emaberto.inep.gov.br/index.php/rbep/article/view/1277/12511>. Acesso em: 18 fev. 2019.

CARDIA, Mirian Lopes. **Mulheres na história: Bertha Lutz**. 2018. Disponível em: <http://www.arquivonacional.gov.br/br/difusao/arquivo-na-historia/908-mulheres-na-historia-bertha-lutz.html>. Acesso em: 8 nov. 2018.

COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR – CAPES. **Qualis**. 2017. Disponível em: <http://www.capes.gov.br/acessoainformacao/perguntas-frequentes/avaliacao-da-pos-graduacao/7422-qualis>. Acesso em: 24 nov. 2018.

CRUZ, Alice. Pode a testemunha modesta falar? A biomedicina em trânsito paradigmático. **Physis Revista de Saúde Coletiva**, Rio de Janeiro, 21 [4]: 1331-1354, 2011. Disponível em: <http://www.scielo.br/pdf/physis/v21n4/a09v21n4.pdf>. Acesso em: 18 ago. 2018.

DAGNINO, Renato. Mais além da participação pública na ciência: buscando uma reorientação dos Estudos sobre Ciência, Tecnologia e Sociedade em Ibero-américa. **Revista Iberoamericana de Ciencia, Tecnología, Sociedad e Innovación**, Número 7 Septiembre - Diciembre 2006. Disponível em <http://www.oei.es/revistactsi/numero7/articulo02.htm>. Acesso em: 8 nov. 2018.

ETZKOWITZ, H. Networks of Innovation: Science, Technology and Development in the Triple Helix Era, **International Journal of Technology Management & Sustainable Development**, 2002.



GOMES, Suely; GOMES, Ronaldo. Ciência e Estado: por uma participação de todos. IN: GOMES, Suely; SANTOS, Andrea Pereira dos; MUELLER, Geisa (orgs). **Letramento informacional**: entendendo a ciência e a comunicação. Pré-print.

FEDERAÇÃO INTERNACIONAL DE ASSOCIAÇÕES DE BIBLIOTECÁRIOS E BIBLIOTECAS. **Manifesto da IFLA sobre a internet**. Haia: IFLA, 2002. Disponível em: <<http://www.ifla.org/>>

FERREIRA, Pedro Lopes. **Estatística descritiva e inferencial**: breves notas. Coimbra: FEUC, 2005. Disponível em: <https://estudogeral.uc.pt/bitstream/10316/9961/1/AP200501.pdf>. Acesso em: 10 fev. 2019.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA - INEP. **Censo da Educação Superior 2017**: notas estatísticas. Brasília: MEC, 2017. Disponível em: http://download.inep.gov.br/educacao_superior/censo_superior/documentos/2018/censo_da_educacao_superior_2017-notas_estatisticas2.pdf. Acesso em: 3 nov. 2018.

LETA, Jaqueline. As mulheres na ciência brasileira: crescimento, contrastes e um perfil de sucesso. **Estudos Avançados**, v 17 (49), 2003.

LOMBARDI, Maria Rosa. Engenheiras e pesquisa científica e tecnológica no Brasil: construção das carreiras e desafios. **Fazendo gênero**, n. 9, 2010. Disponível em: http://www.fazendogenero.ufsc.br/9/resources/anais/1277208903_ARQUIVO_Lombardi_MRFG92010.pdf. Acesso em: 2 mar. 2019.

MARRES, N.; WELTEVREDE, E. Scraping the social? **Journal of cultural economy**, v. 6, n. 3, p. 313-335, 2013. Disponível em: <https://pdfs.semanticscholar.org/8747/2807a9cb672914494cc8ed4b662f6a683454.pdf>. Acesso em: 2 jan. 2019

MARTINS, Gilberto de Andrade, THEÓPHILO, Carlos Renato. **Metodologia da investigação científica para ciências sociais aplicadas**. 2. ed. São Paulo: Atlas, 2009.

NAÇÕES UNIDAS DO BRASIL. **Direitos humanos das mulheres**. 2018. Disponível em: <https://nacoesunidas.org/wp-content/uploads/2018/08/Position-Paper-Direitos-Humanos-das-Mulheres.pdf>. Acesso em: 8 dez. 2018.

OLINTO, Gilda. A inclusão das mulheres nas carreiras de ciência e tecnologia no Brasil. **Inc. Soc.**, Brasília, DF, v. 5 n. 1, p.68-77, jul./dez. 2011.



PORTAL BRASIL. **Mulheres são maioria em universidade e cursos de qualificação.** 2016. Disponível em: < <http://www.brasil.gov.br/economia-e-emprego/2016/03/mulheres-sao-maioria-em-universidades-e-cursos-de-qualificacao>>. Acesso em: 22 jun. 2018.

PARENT IN SCIENCE. **Entendendo a maternidade dentro do universo científico brasileiro.** [S.l.]: Wix.com, 2018. Disponível em: <https://www.parentinscience.com/sobre-o-parent-in-science>. Acesso em: 8 nov. 2018. TENENTE, Renata Cesar Vilardi. *Natureza das profissões.* IN.: BRABO, Tânia Suely Antonelli Marcelino (org.). **Gênero e educação:** lutas do passado, conquistas do presente e perspectivas futuras. São Paulo: Ícone, 2007.

RAGO, Margareth. **A aventura de contar-se:** feminismos, escrita de si e invenções da subjetividade. Campinas, SP: Editora da Unicamp, 2013.

VIEGAS, Daniela, GAZIRE, Marina, REIS, Roberto Alves, BARBOSA, Sílvia Michelle A. Bastos. *Mulheres e contemporaneidade: emergências e urgências.* IN.: MULHERES COMUNICAM: mediações, sociedade e feminismo. Belo Horizonte: Letramento, 2016.

SARDENBERG, Cecília M.B. Da crítica feminista à Ciência a uma Ciência Feminista? In: COSTA, A.A. e SARDENBERG, C.M.B. (orgs.) **Feminismo, Ciência e Tecnologia.** Salvador, Rede Feminista Norte e Nordeste de Estudos e Pesquisa sobre Mulher e Relações de Gênero (REDOR), Núcleo de Estudos Interdisciplinares sobre a Mulher (NEIM), Universidade Federal da Bahia, vol.8 Coleção Bahianas, 2002.

SECRETARIA DE GESTÃO E PLANEJAMENTO - SEGPLAN. **Cenário socioeconômico e ambiental.** 2015. Disponível em: <http://www.sgc.goias.gov.br/upload/arquivos/2015-04/estudo-do-cenario-socioeconomico-e-ambiental.pdf>. Acesso em: 22 jun. 2018.

SCHIENBINGER, Londa. **O feminismo mudou a ciência?** Bauru: EDUSC, 2001.

SILVA, Elizabeth Bortolaia. Des-construindo gênero em ciência e tecnologia. **Cadernos Pagu** (10) 1998: pp.7-20.

WOOLF, Virginia. *Duas mulheres.* In.: WOOLF, Virginia. **Profissões para mulheres e outros artigos femininos.** Porto Alegre: L&PM, 2018.

WORLD ECONOMIC FORUM. **The Global Gender Gap Report.** 2017. Disponível em: http://cite.gov.pt/pt/destaques/complementosDestqs2/WEF_GGGR_2017.pdf. Acesso em: 8 nov. 2018.

Modelo de Visibilidad Académica del Instituto Tecnológico de Costa Rica

***Alexa Ramírez-Vega¹, Patricia Meneses-Guillén², Laura Meneses-Guillén³,
Xiomara Arce-Calderón⁴, Evelyn Solano-Gallardo⁵, Marcela Guzmán-Ovares⁶,
Ivan Mata-Ortega⁷***

¹ Instituto Tecnológico de Costa Rica. Correo: alamirez@tec.ac.cr

² Instituto Tecnológico de Costa Rica. Correo: pmeneses@tec.ac.cr

³ Instituto Tecnológico de Costa Rica. Correo: lmeneses@tec.ac.cr

⁴ Instituto Tecnológico de Costa Rica. Correo: xarce@tec.ac.cr

⁵ Instituto Tecnológico de Costa Rica. Correo: evsolano@tec.ac.cr

⁶ Instituto Tecnológico de Costa Rica. Correo: maguzman@tec.ac.cr

⁷ Instituto Tecnológico de Costa Rica. Correo: imata@tec.ac.cr

Resumen

Actualmente existen mecanismo y buenas prácticas para visibilizar de manera correcta las publicaciones en la web y evitar pérdida de información. A pesar de los esfuerzos realizados por las universidades estas prácticas y mecanismos son desconocidos por los investigadores. Esto hace que muchos publiquen con su firma de autor sin normalizar, escriban de manera equivocada su afiliación institucional, no depositen sus publicaciones en el repositorio institucional, entre otros. Lo anterior no escapa de la realidad del Instituto Tecnológico de Costa Rica (TEC), por lo tanto desde el 2014 se inició con un Modelo de Visibilidad Académica, cuyo objetivo consistió en hacer más visible la producción científica institucional en la web, mediante la identificación de elementos que afectan la visibilidad, lo cual permitió proponer estrategias, generar políticas y conducir los procesos para dar solución a esos problemas. En este trabajo se exponen los resultados obtenidos y las tareas emprendidas con la aplicación del Modelo de Visibilidad del TEC en sus 5 años de funcionamiento.

Palabras clave: Visibilidad web; modelo de visibilidad; estrategias; repositorio institucional; portal de revistas; políticas.

Introducción

En la actualidad las instituciones de enseñanza superior hacen grandes esfuerzos para invertir sus recursos en investigación y extensión, con los años se ha visto el avance en este rubro. Muchos de estos resultados de investigación se divulgan a través de publicaciones de diversos tipos: artículos científicos, ponencias, informes de investigación, tesis, etc. Así mismo, estos documentos son publicados en sitios web que no siempre están accesibles o tienen datos errados,

tanto a nivel de autor como de procedencia de la publicación. En este sentido, las universidades tienen una tarea ardua para identificar todos los trabajos producto de su actividad investigativa.

Dado lo anterior, existen mecanismo y buenas prácticas para visibilizar de manera correcta las publicaciones en la web y evitar pérdida de información [1] [2]. A pesar de los esfuerzos realizados por las universidades estas prácticas y mecanismos son desconocidos por los investigadores. Esto hace que muchos publiquen con su firma de autor sin normalizar [3], escriban de manera equivocada su afiliación institucional, no depositen sus publicaciones en el repositorio institucional, entre otros.

Lo anterior no escapa de la realidad del Instituto Tecnológico de Costa Rica (TEC), por lo tanto desde el 2014 se inició con un Modelo de Visibilidad Académica, cuyo objetivo consistió en hacer más visible la producción científica institucional en la web, mediante la identificación de elementos que afectan la visibilidad, lo cual permitió proponer estrategias, generar políticas y conducir los procesos para dar solución a esos problemas.

Algunas de las principales tareas desarrolladas dentro de este modelo se describen en [4], donde inicialmente se trabajaba de manera independiente y aislada según las necesidades de cada dependencia. Fue así que, con el objetivo de integrar esfuerzos y trabajar de manera institucional en una vía común, se decidió formar de manera extraoficial un grupo de trabajo denominado Grupo de Visibilidad Académica del TEC, el cual le daría seguimiento y desarrollaría las tareas del Modelo de Visibilidad.

En este trabajo se exponen los resultados obtenidos y las tareas emprendidas con la aplicación del Modelo de Visibilidad del TEC en sus 5 años de funcionamiento.

Materiales y metodología

La metodología para la creación de un modelo de Visibilidad Académica del Instituto Tecnológico de Costa Rica fue la siguiente:

- Creación de un grupo interdisciplinar, donde se involucraron representantes de las siguientes entidades: Biblioteca, Editorial, Vicerrectoría de Investigación, Planificación Institucional, Comunicación y Mercadeo.
- Se crea un plan de trabajo conjunto donde cada representante expone los temas sobre visibilidad académica que ha trabajado.
- El plan de trabajo contempla las áreas donde cada dependencia tiene injerencia para trabajar a lo interno y las tareas concretas que se trabajarán en conjunto.
- En el plan se contempla una etapa de detección de problemas de visibilidad con el objetivo de analizar y solventar cada uno de ellos. Cada problema de visibilidad se clasifica según su prioridad (1, 2 o 3) e impacto (alto, medio, bajo).

- Se dividen los problemas encontrados en 3 etapas. La Etapa 1 contempla los problemas de prioridad 1 e impacto alto. La Etapa 2 incluye los problemas de prioridad 2 e impacto alto o medio. Y la Etapa 3 trabaja los problemas de prioridad 3 con impacto medio o bajo.
- Una vez solventado alguno de los problemas de la Etapa 1 se continúa con el siguiente según su categorización. Luego se procede con la Etapa 2 y finalmente con la Etapa 3.

Resultados

En el cuadro 1 se exponen los principales problemas de visibilidad detectados y las tareas emprendidas para solventarlos. Algunas de estas acciones y sus resultados parciales se encuentran en [5].

Cuadro 1. Categorización de problemas de visibilidad por prioridad e impacto

Problema	Prioridad	Impacto
1-Duplicidad de dominios	1	Alto
2-Fortalecimiento de revistas científicas del TEC	2	Medio
3-Fortalecimiento de repositorio institucional	2	Medio
4-Página web institucional	1	Alto
5-Publicaciones científicas en otras revistas	1	Alto
6-Divulgación	3	Medio
7-Capacitación	3	Bajo
8-Sistema de perfiles académicos de investigadores	2	Bajo

Después de identificar los problemas de visibilidad, se procedió a categorizarlos según su prioridad e impacto. De esta manera, se tomaron como punto de partida los problemas de prioridad 1 e impacto alto. De los 8 problemas identificados en el cuadro 1, se tomaron como punto de partida el 1, 4 y 5 que conformaron la etapa 1 del proceso de mejora. Seguidamente, en la etapa 2 se tomaron los problemas 2 y 3. Y finalmente para la etapa 3 se tomaron los problemas 6, 7 y 8.

A continuación se exponen las acciones emprendidas para solventar cada uno de los problemas de visibilidad de la Etapa 1.

Unificación de dominios

Desde su creación el Instituto Tecnológico de Costa Rica utilizaba el dominio `itcr.ac.cr`, posteriormente y para posicionar el nombre de marca TEC se decidió alojar el sitio principal de la institución en `tec.ac.cr`, lo cual generó que la producción estuviera dividida en dos dominios, esto afectó su visibilidad en la web, debido a que se interpretaban ambos como entes o instituciones diferentes.

Para centralizar toda la producción y contenidos del TEC en un mismo dominio, se realizaron varias acciones:

- Análisis de la situación real de la distribución de los documentos en los dos dominios y la visibilidad en la web.
- Limitaciones que tenía la página web en relación con el personal, la plataforma y el equipo.
- El traslado de la responsabilidad de la página web a un ente centralizado,
- Reunión con los encargados de todos los sistemas de información y sitios web del TEC; en esta reunión se define que el dominio donde debían alojarse los subdirectorios era `tec.ac.cr`, lo cual fue ratificado por la Rectoría.
- Coordinación del proceso de migración del dominio `itcr.ac.cr` de los sistemas de información y sitios web a un único dominio.

Como resultado, para febrero del 2015 ya se había migrado aproximadamente el 85% de los sistemas al dominio `tec.ac.cr`. Además, se da seguimiento a los dominios web que se encuentran fuera del dominio `tec.ac.cr`, pero que son creados por funcionarios del TEC para divulgar actividades institucionales y se coordinan estrategias para la migración de esos dominios a subdirectorios dentro de `tec.ac.cr`.

Página web institucional

Dado que la página web institucional es la llave de entrada de los usuarios hacia el contenido académico, esta fue considerada desde el inicio del proceso de mejoramiento de la visibilidad. Así, se desarrollaron las siguientes acciones: un diagnóstico inicial contempló la página web del momento y allí quedaron en evidencia las limitaciones de personal, plataforma y equipo, entre otras; se trasladó al Consejo de Rectoría la preocupación expresada en consenso con los encargados de los sistemas de información y sitios web del TEC, según la cual la web institucional requería ser considerada como un nuevo proyecto integral y debía ser declarada de emergencia institucional.

La Oficina de Comunicación y Mercadeo del TEC ha desarrollado actividades como: rediseño del sitio web; propuesta y aprobación por parte del Consejo de Rectoría del uso de la nueva plataforma, con el sistema de gestión de contenido *Drupal*; construcción del nuevo sitio web; propuesta y aprobación de un modelo de gobernanza que permite a las instancias autogestionar su contenido; propuesta y aprobación de los lineamientos del dominio del sitio web; e inicio del proceso de

repatriación de aquellos sitios que fueron creados fuera del dominio institucional (más de 40 sitios).

Visibilidad de publicaciones científicas del TEC en otras revistas

En este tema se impartió el taller *Google Scholar* y su importancia en la visibilidad de la investigación, en nueve fechas distintas, en dos sedes del TEC. El taller tuvo por objetivo capacitar a los investigadores en el uso de *Google Scholar*, haciendo énfasis en la creación del perfil de citas. Con esto se logró que se tomara conciencia de la importancia de la herramienta, así como pautas para localizar las publicaciones propias en la web. Se logró capacitar a más de 100 investigadores e incrementar exponencialmente la cantidad de perfiles de investigadores del TEC [6].

También se impartieron charlas y se realizaron reuniones y comunicados de prensa con el fin de sensibilizar y estandarizar los nombres de los autores del TEC en las publicaciones, así como el nombre de la afiliación. Se han coordinado acciones para lograr avanzar en la normalización del nombre del TEC (afiliación) en los artículos publicados en revistas indizadas en los principales índices internacionales; se ha dado seguimiento a los perfiles en *Google Scholar* a fin de validar la información que contienen; y la Biblioteca, por medio de talleres dirigidos a investigadores y docentes, los ha motivado para que creen su perfil en ORCID. Todo esto acorde a las tendencias internacionales sobre normalización y creación de firma digital de autor [3].

Después de implementadas las acciones de la Etapa 1, se procedió a realizar las actividades que permitirían mejorar o solucionar por completo los problemas de la Etapa 2.

Fortalecimiento de revistas científicas del TEC

Como se menciona en [7], la principal problemática encontrada en las revistas del TEC fue la escasa presencia de sus contenidos en la web. Para ello se desarrollaron varias acciones: se creó el Portal de Revistas académicas utilizando la plataforma *Open Journal Systems* (OJS), con el fin que los artículos publicados en revistas del TEC estén disponibles en Internet; se trabajó y asesoró a los editores de revistas para incluirlas en el Portal; se avanzó con los editores en el proceso de indización de sus revistas en índices iberoamericanos; se creó un espacio en el Portal de Revistas para albergar memorias de congresos y otros eventos realizados en el TEC; se cosecharon todos los artículos del Portal de Revistas para ponerlos a disposición en el Repositorio Institucional; y se monitorea periódicamente el Portal con la herramienta *Google Analytics*, para constatar el alcance de los artículos que se publican en las revistas del TEC. En la figura 1 se muestra el crecimiento de visitas del sitio del Portal de revistas desde su creación en 2012. Además, se capacita de manera constante a los editores en materia de visibilidad, SEO académico y marketing científico con el objetivo de poner en práctica los mejoras lineamientos para mejorar su visibilidad en la web de contenido académico [8].



Figura 1. Visitas por año al Portal de revistas del TEC

Fortalecimiento de Repositorio Institucional

El Repositorio del TEC fue creado en el 2002 y es administrado por la Biblioteca de esta institución. Sus principales funciones son promover y facilitar la divulgación de la información científica, así como preservar el patrimonio científico y académico de la Institución.

Con el fin de fortalecerlo, los encargados del Repositorio han desarrollado las siguientes acciones [9] [10]:

- Cambio del repositorio del dominio de itcr.ac.cr a tec.ac.cr
- Aplicación de estándares de metadatos DRIVE 0.2 en el Repositorio
- Cambio de la plataforma y de la versión de *DSpace* en la que se desarrolla el Repositorio
- Creación de las políticas del Repositorio
- Estandarizar los metadatos según *OpenAIRE*, lo cual le permite ser cosechado por *Kímuk* (cosechador nacional) y por La Referencia (cosechador latinoamericano).

También, se cosecharon todos los artículos del Portal de Revistas del TEC para ponerlos a disposición en el Repositorio Institucional; se sensibilizó al personal de las unidades académicas para que envíen al Repositorio los documentos académicos y las tesis de graduación; y se modificó la estructura de las comunidades y sub-comunidades del Repositorio, con el fin de adecuarlo a las necesidades de la institución. Además, el sitio del Reposito Institucional es monitoreado mediante *Google Analytics*, en la figura 2 se muestran las visitas y

sesiones recibidas desde 2012. Actualmente se está trabajando en otra estructura para las comunidades de trabajos de graduación.

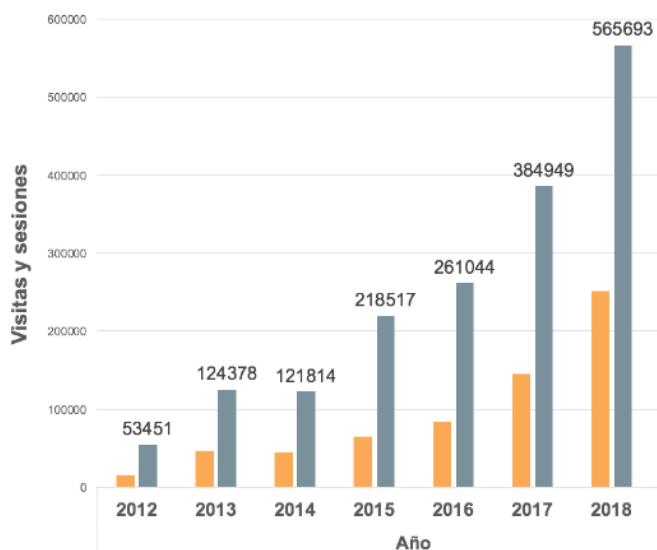


Figura 2. Cantidad de visitas y sesiones por año del Repositorio Institucional

Finalmente, en la Etapa 3 se implementaron acciones, que aún se están realizando para solventar los problemas detectados.

Divulgación

En este campo, se han utilizado estrategias de divulgación para dar a conocer temas como normalización del nombre de autor y afiliación institucional; cambios en el perfil de Google Académico; nuevas métricas, como *Altmetrics*; e identificador ORCID. El objetivo es lograr una amplia participación de las personas involucradas.

Capacitación

Por otra parte, el Grupo de Visibilidad Académica ha recibido capacitaciones que le han permitido actualizarse en el tema de visibilidad de la producción académica. Incluye acceso abierto, repositorios, sistemas de información de investigación, visibilidad de publicaciones científicas, bases de datos de revistas indizadas en índices internacionales, nuevas herramientas para medir la investigación y herramientas para ser utilizadas en web para medir visibilidad, entre otros.

El grupo ha presentado diversas ponencias y pósteres sobre el quehacer del TEC en el campo de la visibilidad en congresos principalmente internacionales y se han publicado artículos en revistas científicas o de divulgación científica.

También se organizó y desarrolló el taller de “Visibilidad académica, rankings, producción científica y nuevas métricas”, efectuado los días 25 y 26 de junio del 2018 en el Hotel Crown Plaza Corobicí. Este evento académico fue dirigido a las universidades nacionales y se brindó un taller individual para cada una de las cinco universidades públicas sobre visibilidad académica y rankings universitarios, con la participación de dos expositores internacionales, en los cuales se contó con la participación de más de 60 funcionarios universitarios.

En enero y febrero del presente año se impartió el Taller de Visibilidad de la producción académica, dirigido a investigadores, extensionistas y gestores de investigación del TEC.

Perfiles académicos de investigadores

Actualmente, el Grupo de Visibilidad se encuentra trabajando en esta etapa con el objetivo de implementar un Sistema de Gestión de la Investigación (o CRIS por sus siglas en inglés). En primer lugar, se evaluaron algunas iniciativas implementadas en la institución relacionadas con perfiles académicos de investigadores. Estas iniciativas, aunque son un buen punto de partida, presentan muchas carencias en cuanto a escalabilidad, ingesta de información, interoperabilidad, entre otras. De esta manera, y siguiendo con los lineamientos nacionales e internacionales se están estudiando diversas alternativas viables para gestionar de manera integral los perfiles de investigadores de la institución [11] [12].

Una vez finalizada la Etapa 3, se entra en un proceso de reevaluación y análisis para identificar nuevos puntos de mejora en temas de visibilidad académica.

Conclusiones

Las acciones llevadas a cabo han permitido aumentar la visibilidad de la producción científica del TEC. Ejemplo de ellos son: la unificación de dominios del TEC en un solo dominio –tec.ac.cr-; una mayor visibilidad y utilización del Repositorio Institucional y del Portal de Revistas TEC, esto se evidenció en el constante aumento en las visitas recibidas en ambos sitios entre 2012 y 2018. Además, el haber concebido la página web institucional como un proyecto integral, que implicó el rediseño de la página, mejorando en calidad y cantidad de contenidos disponibles en ese sitio. Así mismo, se logró un mayor número de académicos con perfil en *Google Scholar* y diversas publicaciones sobre el tema de visibilidad académica.

También es importante mencionar la mejora en los rankings internacionales, específicamente *Webometrics*. El TEC ha pasado de la posición 10 a nivel nacional y 6297 a nivel internacional en el 2014, a la posición 4 a nivel nacional y 2792 a nivel internacional en 2018. Cabe mencionar que esto es solo un indicador más que permite conocer el estado de la visibilidad académica TEC en la web, según estos criterios específicos. Además, estas posiciones son muy sensibles a fluctuar debido a los cambios constantes de metodología y mejora de otras universidades.

Como trabajo futuro se espera seguir implementado este Modelo de Visibilidad Académica con el objetivo de detectar otros problemas (o puntos de mejora) de visibilidad que aún no se han solventado del todo. Además, cabe mencionar que aunque el trabajo se ha realizado de manera conjunta y coordinada el grupo de trabajo involucrado no cuenta con una figura oficial dentro de la institución, razón por la cual se está trabajando en un plan que incluya entre sus prioridades la creación de una entidad encargada de estos temas de manera oficial. Esto, además de brindar oficialidad a las acciones, permite obtener financiamiento y apoyos tanto interno como externos.

Referencias

- [1] M. García Raga y L. A. Algas Hechavarría, «Creación de perfiles académicos para incrementar la visibilidad de la ciencia», *Multimed*, vol. 21, n.º 2, pp. 1-3, jun. 2017.
- [2] F.-A. López, «Visibilidad e impacto de los repositorios digitales en acceso abierto», *Bibl. Bibl. Bol. Electrónico ABGRA*, n.º 5, 2013.
- [3] L. M. Perilla-Rodríguez y A. M. Pérez-Acosta, «Normalización de las firmas de autor y de institución: de los nombres a los ID», *Av. En Psicol. Latinoam.*, vol. 34, n.º 3, pp. 437-439, dic. 2016.
- [4] A. Ramírez-Vega, «Programa de difusión de la producción intelectual del Instituto Tecnológico de Costa Rica».
- [5] A. Ramírez-Vega, «Visibilidad de la producción científica y académica del Instituto Tecnológico de Costa Rica», *Investiga TEC*, n.º 24, p. página-13, 2015.
- [6] A. Ramírez-Vega y P. Meneses-Guillén, «Google Scholar y su importancia en la visibilidad de la investigación del Instituto Tecnológico de Costa Rica», *Investiga.TEC*, vol. 0, n.º 28, p. 5, feb. 2017.
- [7] A. Ramírez-Vega, «Desarrollo e impacto del portal de revistas electrónicas del TEC», *Investiga.TEC*, n.º 18, p. pág. 16-19, oct. 2013.
- [8] A. Ramírez-Vega, «Uso de marketing digital para la divulgación y visibilidad de publicaciones académicas en acceso abierto».
- [9] I. Alpizar-Chacon, X. Arce-Calderón, y L. Meneses-Guillén, «Experiencia de fortalecimiento del RepositorioTEC: mejoras en el proceso de recolección de documentos para aumentar la visibilidad», 2017.
- [10] I. Alpizar-Chacón, M. Chacón-Rivas, y L. Meneses-Guillén, «Experiencia en el uso de herramientas y procesos para mejorar la visibilidad de un repositorio institucional basado en DSpace: Repositorio TEC», 2015.
- [11] A. García-Holgado, J.-A. Merlo-Vega, y F. García-Peñalvo, «Diseño de un plan de visibilidad científica e identidad digital para los investigadores de la Universidad de Guadalajara (México)», *Ibersid Rev. Sist. Inf. Doc.*, vol. 11, n.º 1, pp. 83-92, 2017.
- [12] Á. Rodríguez Terán, «Sistemas de Gestión de la Investigación: aproximación a los CRIS Institucionales.», PhD Thesis, Universidad de Salamanca, 2015.

DIA 3

CIÊNCIA ABERTA, REPOSITÓRIOS DE DADOS DE PESQUISA, PLANO DE GESTÃO DE DADOS

Práticas e políticas

Ética e Direitos autorais

Gestão e curadoria

Interoperabilidade

Visibilidade e promoção de dados abertos na pesquisa

Preservação digital

Reprodutibilidade da investigação na ciência

A integração do Arca - Repositório Institucional da Fiocruz com a Plataforma de Ciência de Dados aplicada à Saúde

Claudete Fernandes de Queiroz¹, Ana Maria Neves Maranhão², Luciana Danielli de Araujo³, Andrea F. Gonçalves do Nascimento⁴, Raphael Belchior Rodrigues⁵, Éder de Almeida Freyre⁶, Jefferson da Costa Lima⁷, Marcel de Moraes Pedroso⁸

¹ Fundação Oswaldo Cruz, Instituto de Comunicação e Informação Científica e Tecnologia em Saúde, Rio de Janeiro, Brasil.
Email: claudete.queiroz@icict.fiocruz.br

² Fundação Oswaldo Cruz, Instituto de Comunicação e Informação Científica e Tecnologia em Saúde, Rio de Janeiro, Brasil. Email: anamaranhao01@gmail.com

³ Fundação Oswaldo Cruz, Instituto de Comunicação e Informação Científica e Tecnologia em Saúde, Rio de Janeiro, Brasil.
Email: luciana.danielli@icict.fiocruz.br

⁴ Fundação Oswaldo Cruz, Instituto de Comunicação e Informação Científica e Tecnologia em Saúde, Rio de Janeiro, Brasil.
Email: andrea.goncalves@icict.fiocruz.br

⁵ Fundação Oswaldo Cruz, Instituto de Comunicação e Informação Científica e Tecnologia em Saúde, Rio de Janeiro, Brasil.
Email: raphael.rodriques@icict.fiocruz.br

⁶ Fundação Oswaldo Cruz, Instituto de Comunicação e Informação Científica e Tecnologia em Saúde, Rio de Janeiro, Brasil. Email: eder.freyre@icict.fiocruz.br

⁷ Fundação Oswaldo Cruz, Instituto de Comunicação e Informação Científica e Tecnologia em Saúde, Rio de Janeiro, Brasil. Email: jefferson.lima@icict.fiocruz.br

⁸ Fundação Oswaldo Cruz, Instituto de Comunicação e Informação Científica e Tecnologia em Saúde, Rio de Janeiro, Brasil.
Email: marcel.pedroso@icict.fiocruz.br

Resumo

Apresenta o projeto desenvolvido entre o Laboratório de Ciência de Dados aplicada à Saúde, do Instituto de Informação Científica e Tecnológica em Saúde (ICICT) e o Arca – Repositório Institucional da Fiocruz. O projeto teve como objetivos: melhorar a curadoria dos dados inseridos no repositório institucional, visando a qualidade das informações, e a recuperação e a visualização de dados, oferecendo uma plataforma que permite a extração de informações com potencial de uso pela gestão e pela pesquisa. No processo de curadoria foi possível identificar inconsistências no preenchimento dos metadados, utilizando classificação automática e *machine learning*, e consequente correção, de forma a garantir a qualidade das informações e dos dados extraídos. Outro fator importante para a realização do projeto foi a utilização do software Kibana e do Elasticsearch para a visualização de dados de forma dinâmica, oferecendo uma

plataforma de exploração interativa para extração e mineração de dados. O software permitiu a utilização de filtros e combinações de dados contidos no Arca, como produção por tipo de material, Unidades da Fiocruz, assunto, autor, ano e direito autoral de forma que possam ser manipulados pelas diferentes unidades/comunidades representadas no Repositório Institucional.

Palavras-chave: Arca - Repositório Institucional da Fiocruz. Ciência de Dados. Visualização de Dados. Curadoria Digital.

Abstract

It presents the project developed between the Data Science Laboratory applied to Health, the Institute of Scientific and Technological Information in Health (ICT) and the Arca - Institutional Repository of Fiocruz. The objective of the project was to improve the curation of data inserted in the institutional repository, aiming at the quality of information, and the retrieval and visualization of data, offering a platform that allows the extraction of information with potential use by management and research. In the curatorial process it was possible to identify inconsistencies in the metadata filling, using automatic classification and machine learning, and consequent correction, in order to guarantee the quality of information and data extracted. Another important factor for the realization of the project was the use of Kibana and Elasticsearch software to dynamically display data, offering an interactive exploration platform for data mining and extraction. The software allowed the use of filters and combinations of data contained in the Ark, such as production by material type, Fiocruz Units, subject, author, year and copyright so that they can be manipulated by the different units / communities represented in the Institutional Repository.

Keywords: Arca - Institutional Repository of Fiocruz. Data Science; Data Visualization. Digital Curatorship.

Introdução

Ciência de Dados é um campo que objetiva reunir um conjunto de estratégias, ferramentas e técnicas que combina métodos tradicionais de análise com algoritmos sofisticados para processar grandes volumes de dados em formatos diversos - dados estruturados, semiestruturados e não estruturados. Esse processo de análise, no âmbito da Ciência de Dados, envolve fases como coleta e ingestão; pré-processamento; análise exploratória; mineração de dados; e pós-processamento (PEDROSO, 2017).

Segundo Sayão e Sales (2015)

o reconhecimento do potencial informacional dos dados de pesquisa para a ciência contemporânea transformou a visão que os caracterizava como simples subprodutos dos processos de pesquisa. Atualmente os pesquisadores, as instituições acadêmicas e as agências de fomento à pesquisa começam a compreender que esses dados, se devidamente tratados, preservados e gerenciados, podem constituir uma fonte inestimável de recursos informacionais para a pesquisa científica e para o ensino da ciência.

Para além do desenvolvimento de pesquisas de alta qualidade e excelência, a gestão eficiente de dados permite ações de curadoria em repositórios institucionais, por exemplo, aumentando a qualidade e a confiabilidade dos registros depositados.

Na área de gestão do conhecimento em instituições de ensino e pesquisa, os Repositórios Institucionais são importantes ferramentas de gestão e não só de armazenamento e disseminação, pois permitem, com a utilização de técnicas e ferramentas adequadas, a recuperação e visualização dos dados ali contidos de forma dinâmica e objetiva, agregando-se imensurável valor às funções dos RIs.

O Repositório Institucional da Fiocruz – Arca tem como objetivo reunir e disponibilizar em um único local a produção intelectual produzida na Instituição e foi estabelecido como principal instrumento para realização da política de acesso aberto institucional. A Política de Acesso Aberto ao Conhecimento¹, implantada em 2014, estabelece como mandatório o depósito de teses, dissertações e artigos publicados, promovendo, também, desta forma, o alinhamento da Fiocruz com o movimento internacional de acesso aberto.

O Repositório atua, ainda, como uma rica fonte de informação para o desenvolvimento de novas linhas de pesquisas em todas as áreas da saúde pública, corroborando assim, com a missão da Fiocruz perante a sociedade:

Produzir, disseminar e compartilhar conhecimentos e tecnologias voltados para o fortalecimento e a consolidação do Sistema Único de Saúde (SUS) e que contribuam para a promoção da saúde e da qualidade de vida da população brasileira, para a redução das desigualdades sociais e para a dinâmica nacional de inovação, tendo a defesa do direito à saúde e da cidadania ampla como valores centrais (FUNDAÇÃO OSWALDO CRUZ, 2019).

O Repositório foi criado em 2007, sendo lançado oficialmente como institucional em 2011. Está organizado em Comunidades que correspondem às Unidades Técnico-Científicas da Fiocruz e é mantido pelo Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (ICICT).

O RI utiliza o software livre DSpace² e sua alimentação é descentralizada, realizada pelas diversas unidades e seus respectivos setores e atores: bibliotecas, pesquisadores, etc., cabendo ao ICICT e a equipe técnica do Arca a gestão central.

Após o estabelecimento da Política de Acesso Aberto, o RI tem apresentado crescimento exponencial no número de depósitos realizados, em torno de 30%, conforme Tabela 1.

¹ https://portal.fiocruz.br/sites/portal.fiocruz.br/files/documentos/portaria_-_politica_de_acesso_aberto_ao_conhecimento_na_fiocruz.pdf

² No momento, estamos utilizando a versão 4.7, mas realizando testes para mudar para a versão 6.3.

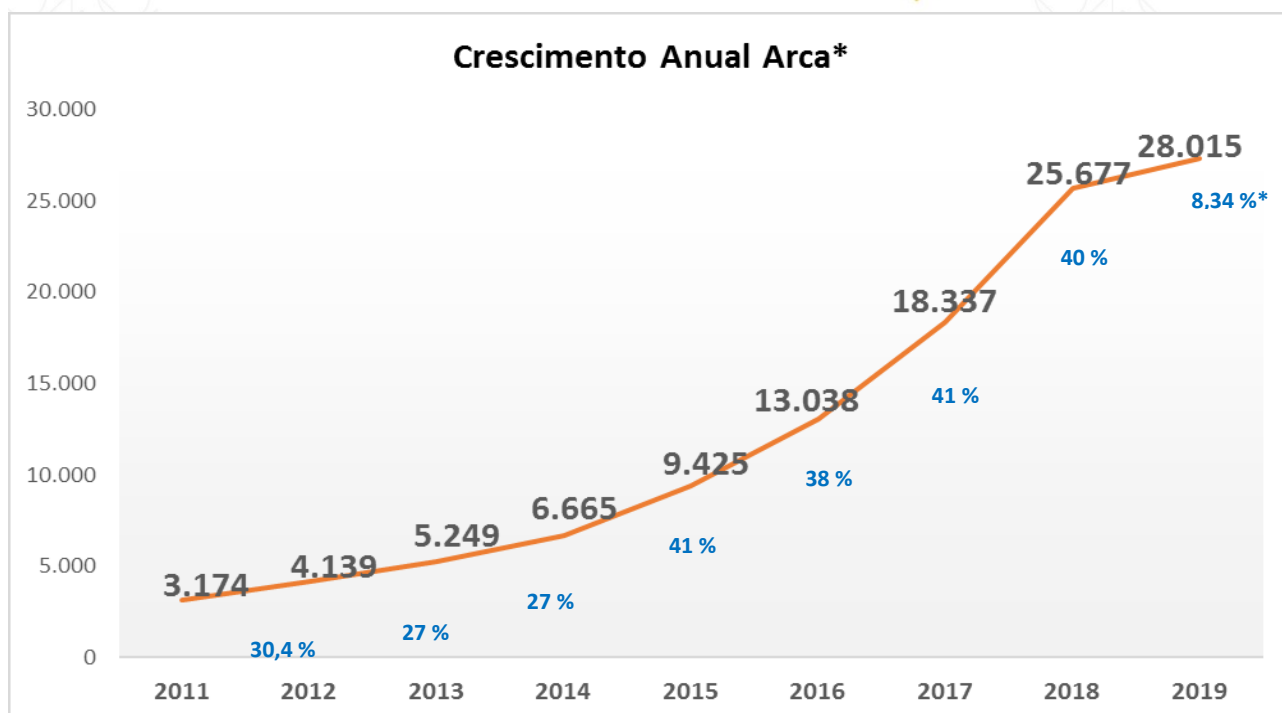


Tabela 1: Crescimento Anual Arca (Atualizado até abril 2019).

Fonte: Fiocruz

No campo das Ciências de Dados, o ICICT, unidade técnico-científica que atua nas áreas de comunicação e informação para a saúde, na Fiocruz, vem desenvolvendo uma série de pesquisas e estudos, como o desenvolvido pelo Laboratório de Informação em Saúde (LIS), que criou e disponibiliza uma Plataforma para processamento de grandes volumes de dados.

A Plataforma de Ciência de Dados aplicada à Saúde da Fiocruz é fruto de:

projeto de pesquisa e desenvolvimento tecnológico do Laboratório de Informação em Saúde do Instituto de Comunicação e Informação Científica e Tecnológica em Saúde da Fundação Oswaldo Cruz (Lis/Icict/Fiocruz), em parceria com o Laboratório Nacional de Computação Científica (LNCC), que disponibiliza para a comunidade científica e gestores um serviço online de armazenamento, gestão e análise de dados em saúde, possibilitando o uso de estratégias como análise visual, mineração de dados, big data, aprendizagem de máquina, dentre outras³ (FUNDAÇÃO OSWALDO CRUZ, 2019).

Dentro deste contexto, a principal proposta da plataforma é “coletar, processar e analisar informações por meio da Ciência de Dados que permitirá planejar, monitorar e avaliar políticas públicas e serviços de saúde em tempo real, gerando indicadores de alerta e painéis de monitoramento bastante específicos”⁴ (FUNDAÇÃO OSWALDO CRUZ, 2019).

³ <https://bigdata.icict.fiocruz.br/>

⁴ A Plataforma irá proporcionar aos pesquisadores, docentes e discentes do Icict e de outras instituições acesso facilitado e qualificado a grandes quantidades de microdados.

Trabalhar com um grande volume de informação requer habilidades e técnicas que se destacam pela capacidade de gerenciar grandes ou complexos sistemas, promovendo a qualidade das informações, consistência dos metadados, interação e integração entre bases de dados.

Desta forma, percebeu-se a importância na formatação de uma parceria entre o RI-Arca e a Plataforma de Ciência de Dados Aplicada à Saúde, para definir estratégias e ferramentas que auxiliassem na coleta, transformação e análise dos dados disponibilizados no Repositório.

Essa parceria resultou no projeto “Ciência de Dados aplicada ao Arca” (Figura 1), com os seguintes objetivos:

- ✓ Curadoria de dados: identificar inconsistências no preenchimento dos metadados do Arca, por meio da classificação automática utilizando *machine learning*, e consequente correção, visando qualidade das informações e dos dados extraídos, facilitando o trabalho de curadoria iniciado em 2015;
- ✓ Recuperação da informação e visualização de dados: oferecer uma plataforma de exploração interativa para visualização e extração de dados, utilizando filtros e combinações de dados contidos no Arca, e que possam ser manipulados pelas diferentes unidades representadas no Repositório Institucional.



Figura 1: Fonte: <https://bigdata.icict.fiocruz.br/ciencia-de-dados-aplicada-ao-arca>

Metodologia

A metodologia proposta envolveu a formalização da parceria e desenvolvimento do Projeto entre o Laboratório de Ciência de Dados e o Arca – Repositório Institucional da Fiocruz, visando o estabelecimento de critérios e procedimentos que atendessem a demanda pelo gerenciamento e visualização de dados contidos no RI.

Foram desenvolvidas as seguintes etapas/atividades:

- estabelecimento das áreas que comporiam a página de visualização de dados, *dashboard*, do RI Arca, como: ano de publicação, assunto, unidade/comunidade, tipologia, autor e direito autoral;
- extração de todos os registros da base de dados, em formato xml, referente as coleções de teses e dissertações, dos programas de pós-graduação da Fiocruz, e dos artigos científicos publicados, tipologias mandatórias da Política de Acesso Aberto ao Conhecimento da Instituição⁵;
- após a extração dos registros, estabelecimento de critérios visando a melhor visualização das informações, reunindo variantes das palavras – plural e singular, sinônimos e homônimos, e o corte em um número de frequência mínimo, visando reunir num universo delimitado os assuntos que apareciam com maior frequência no RI Arca;
- identificação de inconsistências no preenchimento de metadados, como, por exemplo, registros com mais de uma URI;
- emissão e envio de relatórios aos responsáveis pela alimentação nas diversas unidades institucionais para que realizassem as correções necessárias; (Figura 2);
- realizadas as correções, realização de nova exportação, seguindo os mesmos critérios para a verificação dos acertos descritos.

Cabe ressaltar que este último procedimento se tornou sistemático e foi incorporado as tarefas rotineiras, tanto para a equipe do RI Arca, quanto para as Comunidades que compõem o RI, promovendo assim, um trabalho mais ágil e cooperativo em Rede, com o estabelecimento de estratégias de coleta, gestão e correção dos metadados.

Repositório Institucional Arca
Casa de Oswaldo Cruz
Identificação de erros nos metadados na coleção Artigos de Periódicos
Anos verificados 2010 a 2019

Identificador do Documento (handle)	Título do documento	Campo do metadado com problemas	Problema detectado	Solução
http://www.arca.fiocruz.br/handle/icict/12101	La "cultura de la sobrevivencia" y la salud pública internacional en América Latina: la Guerra Fria y la erradicación de enfermedades a mediados del siglo XX	dc.subject.en	As palavras-chaves estão repetidas 4 vezes: international health; Cold War; Latin America; malaria; eradication	Remover as palavras repetidas
		dc.subject.es	As palavras-chaves estão sem padrão começando algumas com maiúscula e outras com minúscula	Padronizar as palavras-chaves, conforme descrito no manual de preenchimento de metadados
		dc.date.issue	Falta de padronização na data 15-Jan-2015	Corrigir para 2015
http://www.arca.fiocruz.br/handle/icict/10740	Uma abordagem arquivística: os documentos de um laboratório das ciências biomédicas	dc.identifier.citation	Falta de padronização v.19, n.1, jan-mar. 2012, p.303-323	Corrigir para v. 19, n. 1, p. 303-323, jan./mar. 2012
http://www.arca.fiocruz.br/handle/icict/1474	Gestão do Conhecimento: ainda um obscuro objeto de desejo?	dc.identifier.citation	Falta de padronização SANTOS, P.X.; REIS, M.E.A. Gestão do conhecimento: ainda um obscuro objeto de desejo? RECIIS: Revista Eletrônica de Comunicação, Informação & Inovação em Saúde, Rio de Janeiro, v. 16, n. 1, p. 1-10, 2012.	Corrigir para SANTOS, Paula Xavier; REIS, Maria Elisa Andriea dos. Gestão do conhecimento: ainda um obscuro objeto de desejo? RECIIS: Revista Eletrônica de Comunicação, Informação & Inovação em Saúde, Rio de Janeiro, v. 16, n. 1, p. 1-10, 2012.

Figura 2: Exemplo de Relatório após finalização da curadoria

⁵ https://portal.fiocruz.br/sites/portal.fiocruz.br/files/documentos/portaria_-_politica_de_acesso_aberto_ao_conhecimento_na_fiocruz.pdf

Com a finalização dessas etapas, foi disponibilizada uma página no Repositório para a visualização dos dados gerais extraídos, através de um *dashboard* com os metadados definidos (ano de publicação, assunto, unidade/comunidade, tipologia, autor e direito autoral), conforme apresentado na Figura 3.

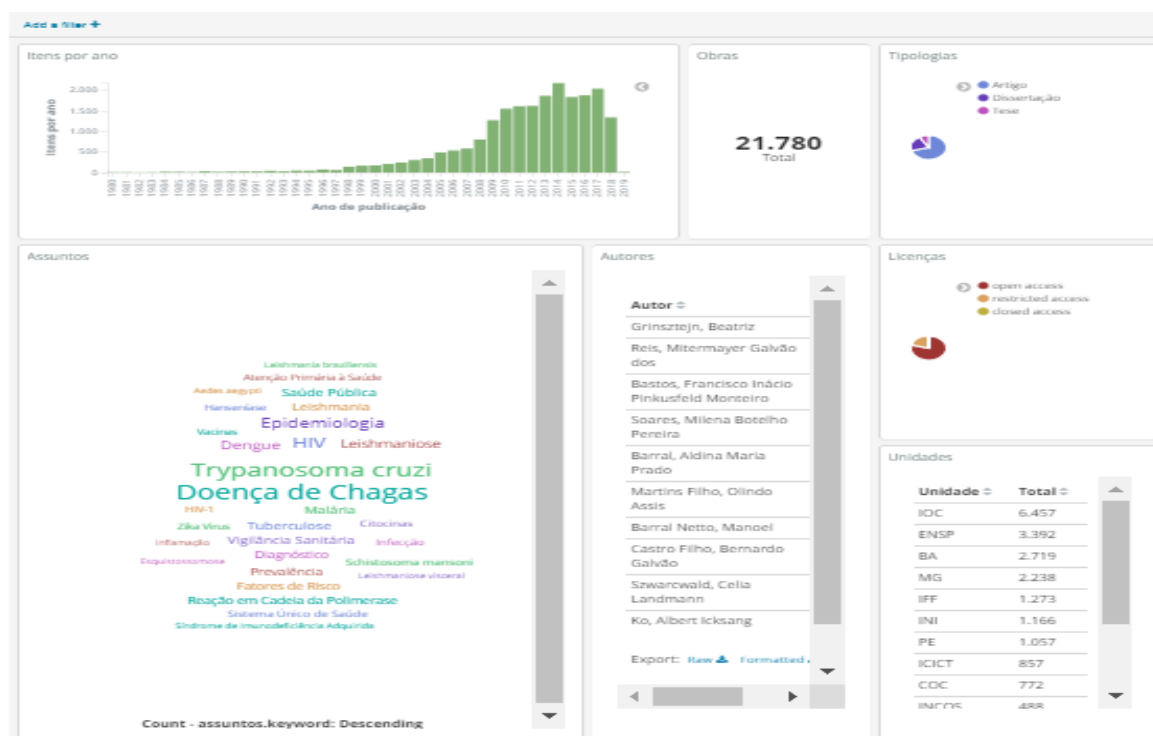


Figura 3: Dashboard com dados gerais do RI - Arca (abril 2019)

Na Figura 3 é possível visualizar as informações de forma dinâmica, permitindo uma visão geral da produção Institucional, como também selecionar uma Comunidade e visualizar quantos documentos foram publicados sobre determinado assunto, em um determinado ano.

No projeto foram utilizados os softwares Elasticsearch e Kibana - o primeiro trabalha com grandes volumes de dados e fornece uma API para a realização de análises dos dados recuperados, e o segundo é um *plugin*, que fornece recursos de visualização para os conteúdos indexados.

Problema

Tendo em vista que a alimentação no RI Arca é descentralizada, sendo realizada através das diversas Unidades e de suas Bibliotecas, além do recurso de autoarquivamento, se tornou fundamental o monitoramento da qualidade dos dados preenchidos através da curadoria digital.

Dentro deste contexto, em 2015, estabeleceu-se um plano de ação para dar início ao trabalho de curadoria digital no RI Arca que tinha como principal objetivo firmar padrões visando a organização das informações e dos objetos digitais dentro do RI (MARANHÃO; QUEIROZ; BELCHIOR, 2017).

O crescimento exponencial no número de depósitos, notadamente, após o estabelecimento da Política de Acesso Aberto ao Conhecimento no ano de 2014, em torno de 160% (período 2014-2018), tornou necessário e fundamental a utilização de mecanismos que facilitassem a curadoria digital, a recuperação e a visualização do conteúdo disponibilizado, permitindo assim, obter um panorama da produção científica institucional, tendo em vista que os RIs são, também, instrumentos de gestão.

Justificativa

A parceria firmada entre o RI Arca e a equipe do Laboratório de Ciência de Dados ajudou a complementar uma lacuna que existia no que se refere a curadoria digital e na gestão dos registros disponibilizados. Além disso, foi possível abordar de forma prática grandes quantidades de dados em diferentes formatos por meio de estratégias e técnicas relacionadas a Ciência de Dados.

Com o estabelecimento de diretrizes e procedimentos, foi viável criar uma interface amigável para a visualização dos dados contidos no RI Arca (Figura 4).

Outro fator importante para a realização do projeto foi que seria possível estabelecer estratégias para a coleta, gestão e correção dos metadados descritos nas diferentes tipologias, oferecendo aos gestores das Comunidades do Repositório, ferramentas que facilitassem e agilisassem os acertos em um tempo mais viável.

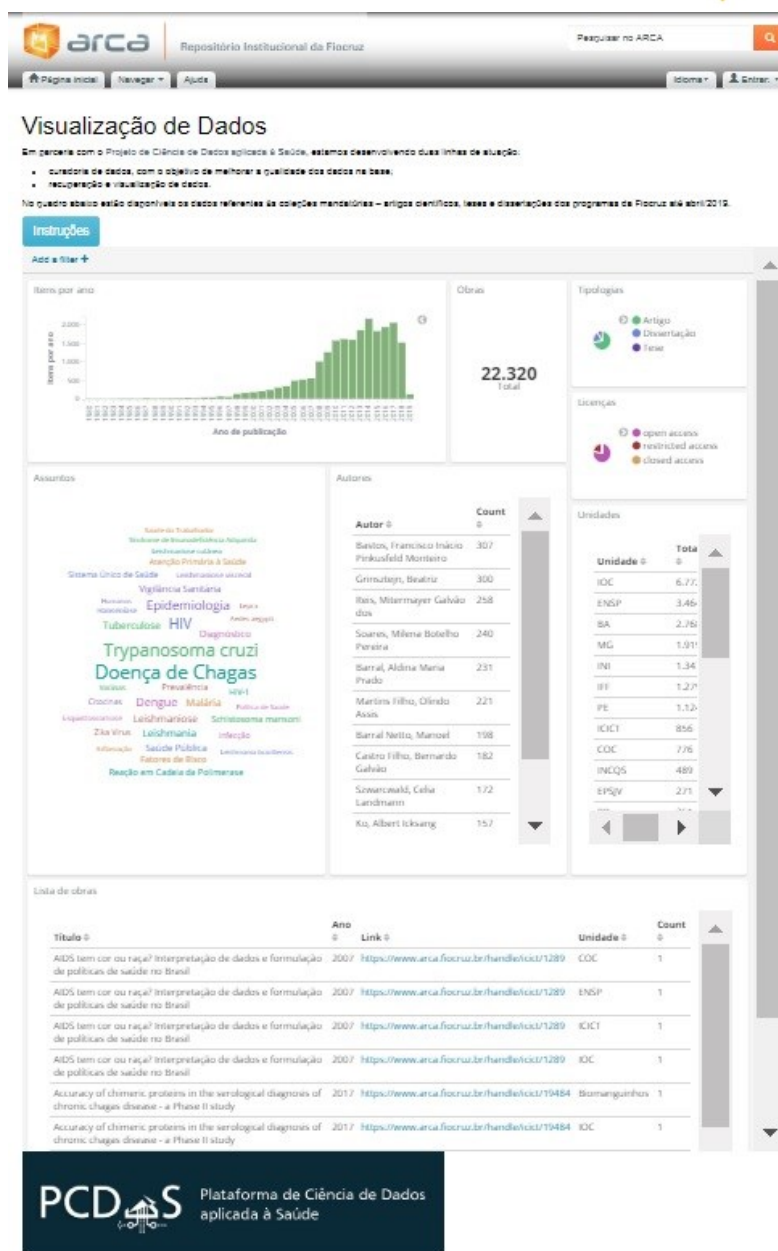


Figura 4: Visualização de Dados no Arca (abril 2019)

Resultados e Discussões

Implantação de uma rotina sistemática no trabalho de curadoria dos dados no RI Arca, de forma que os gestores das Unidades técnico científicas da Fiocruz pudessem visualizar as informações a partir da extração dos registros relevantes. Também foi possível identificar as inconsistências no preenchimento dos metadados, utilizando os sistemas Kibana e Elasticsearch para a classificação automática e correção dos dados, de forma padronizada.

O sistema também possibilitou apresentar uma nuvem de tags com os assuntos mais indexados no Repositório⁶, destacando assim, a importância da indexação e do papel do Bibliotecário na gestão das informações relevantes para o campo da Saúde e Pesquisa dentro da Fiocruz (Figura 5).



Figura 5: Nuvem de Tags – Visualização de Dados
(Assuntos mais Indexados no RI - Arca (abril 2019))

Podemos afirmar, portanto, que a aplicação do Projeto de Ciência de Dados pode propiciar, de forma colaborativa, a melhoria na qualidade dos metadados armazenados, promover a visualização de uma quantidade significativa de informações e garantir a recuperação mais precisa para os usuários do RI.

Outro fator importante para a implantação do projeto foi que através da integração dos sistemas DSpace, Kibana e Elasticsearch, foi possível oferecer soluções de análise, mineração e visualização de dados, através da extração de uma considerável quantidade de registros, proporcionando assim, o acesso aos resultados estruturados e quantificados numa interface mais amigável para o usuário final.

Referências

FUNDAÇÃO OSWALDO CRUZ. **Ciência de Dados aplicada à Saúde**. Rio de Janeiro, 2019. Disponível em:

⁶ É importante lembrar que RI Arca não reproduz necessariamente o que a Fiocruz produz, mas sim o que está depositado.



<<https://bigdata.iciict.fiocruz.br/Apresenta%C3%A7%C3%A3o>>. Acesso em 20 mar. 2019.

FUNDAÇÃO OSWALDO CRUZ. **Perfil institucional**. Rio de Janeiro, 2019. Disponível em: <<https://portal.fiocruz.br/perfil-institucional>>. Acesso em 20 mar. 2019.

FUNDAÇÃO OSWALDO CRUZ. **Sobre o Arca**. Rio de Janeiro, 2019. Disponível em: <<https://www.arca.fiocruz.br/terms/sobre.jsp>>. Acesso em 10 abr. 2019.

MARANHÃO, Ana Maria Neves; DE QUEIROZ, Claudete Fernandes; RODRIGUES, Raphael Belchior. Curadoria Digital de Dados no Arca - Repositório Institucional da Fiocruz: relato de experiência. **RECIIS - Revista Eletrônica de Comunicação, Informação & Inovação em Saúde**, Rio de Janeiro, v. 11, p. 1-4, nov. 2017. Suplemento. Disponível em: <<https://www.arca.fiocruz.br/handle/iciict/23725>>. Acesso em: 02 abr. 2019.

PEDROSO, Marcel de Moraes; LIMA, Jefferson da Costa; ASSEF NETO, Vinicius Belchior. Ciência de Dados aplicada ao Arca: desenvolvimento e disponibilização de ferramentas para recuperação da informação no Repositório Institucional da Fundação Oswaldo Cruz. **RECIIS - Revista Eletrônica de Comunicação, Informação & Inovação em Saúde**, Rio de Janeiro, v. 11, p. 1-5, nov. 2017. Suplemento. Disponível em: <<https://www.arca.fiocruz.br/handle/iciict/23717>>. Acesso em: 02 abr. 2019.

SAYÃO, Luis Fernando; SALES, Luana Farias. **Guia de Gestão de Dados de Pesquisa para Bibliotecários e Pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015. 90 p.

Repositório bilíngue e a disseminação de conhecimento científico e cultural para e sobre educação de surdos

Tania Chalhub¹

¹ PhD. University of Minnesota, Pós-Doutoramento - Instituto Brasileiro de Informação em Ciência e Tecnologia. Professora Adjunta do Instituto Nacional de Educação de Surdos. Coordenadora do Repositório Digital Huet. Email: chalhubtania@gmail.com

Resumo

As tecnologias da informação e comunicação (TIC) impulsionaram a criação de novos espaços de agregação, organização e compartilhamento de informações, contribuindo para que a comunicação se transformasse num bem mais acessível, impactando a educação e a pesquisa. Neste novo cenário a acessibilidade informacional dos surdos foi potencializada significativamente considerando que sua comunicação se estabelece no campo da visualidade. A Língua Brasileira de Sinais (Libras) se concretizou enquanto língua oficial dos surdos pela Lei nº 10.436 de 2002, regulamentada pelo Decreto nº 5.626, 2005, que estabelece a Libras como língua de comunicação dos surdos brasileiros e como canal de acesso à educação. Esta foi uma das maiores conquistas para educação de surdos, que em conjunto com os avanços tecnológicos possibilitou a ampliação da comunicação de conteúdo acadêmico em língua de sinais. O objetivo principal é discutir a disseminação da informação e conhecimento sobre e para educação de surdos via repositório bilíngue (Português e Libras). Esta é uma pesquisa descritiva com abordagem quantitativa e qualitativa, realizada por meio de levantamento da estatística de acesso do Repositório Digital Huet visando a identificar os materiais em Libras e sobre educação de surdos. Projetado e desenvolvido pelo INES, o Repositório foi inaugurado em novembro de 2017 e atualmente conta com mais de 600 objetos de diversos formatos e com mais de quatro milhões de acessos. Dentre os materiais com maior número de acessos estão arquivos fotográficos históricos e vídeos de documentos traduzidos para Libras, vídeos com conteúdo de diferentes áreas do conhecimento como Física e Química e Literatura Infantil.

Palavras-chave: Repositório bilíngue. Acessibilidade informacional. Educação de surdos.

Abstract

Information and communication technologies (ICT) have led to the creation of new spaces for the aggregation, organization and sharing of information, helping to make communication more accessible, impacting education and research. In this new scenario the information accessibility of the deaf was significantly enhanced considering that their communication is established in the field of visibility. The Brazilian Language of Signals (Libras) became an official language of the deaf, Law No. 10,436 of April 24, 2002, regulated by Decree No. 5,626, 2005, recognizing the language of communication of the deaf in Brazil and the right of access to the education in this visual language. Deaf education benefited from advances in technology, especially in this century with the possibility of communicating academic content in sign language. The main objective is to discuss the dissemination of information and knowledge about and for the education of the deaf through the bilingual repository (Portuguese and Libras). Descriptive research with a quantitative and qualitative approach, carried out by means of a survey of the access statistics of the Digital Repository Huet aiming to identify the materials in Libras and about the education of the deaf. Designed and developed by INES, the Repository was inaugurated in November 2017 and currently has more than 600 objects of various formats and with more than four million hits. Among the materials with the greatest number of accesses are historical photographic archives and videos of documents translated into Libras, videos with content from different areas of knowledge such as Physics and Chemistry and Children's Literature.

Keywords: Bilingual repository. Informational accessibility. Education of the deaf

Introdução

A educação de surdos como um tema que merece abordagem diferenciada parece ser fenômeno recente, mas já é pauta de debate no meio educacional há alguns séculos. Porém, a relação deste fenômeno com as tecnologias é realmente uma questão que se faz presente nas últimas décadas com o desenvolvimento de tecnologias assistivas.

Apesar de presente desde o século XVI com educadores como Ponce de Léon, na Espanha, que buscavam ensinar os surdos a ler e a falar, apenas no século XVIII a educação de surdos foi institucionalizada por Abade L'Epée na França, em 1755. Na escola de Paris ele desenvolveu um sistema de comunicação com surdos que ficou conhecido como Sinais Metódicos que deu origem à Língua de Sinais Francesa (LSF). Esta abordagem foi reconhecida como mais eficiente para o ensino de surdos e influenciou o crescimento de escolas específicas para surdos na Europa e em outros países como Estados Unidos da América do Norte e Brasil, para onde migraram alguns professores surdos que haviam experienciado este método no

Instituto Nacional de Jovens Surdos (escola fundada por L'Epée em 1755 e transformada em instituto nacional em 1791) (SACKS, 2015).

Segundo o linguista norte-americano Sacks:

ser deficiente na linguagem é uma das calamidades mais terríveis, porque é apenas por meio da língua que entramos plenamente em nosso estado e cultura Humanos, que nos comunicamos livremente com nossos semelhantes, adquirimos e compartilhamos informações (2015, p. 19).

Para os surdos a língua de sinais é a língua natural de comunicação, sendo uma língua muito recente e até pouco tempo vista como uma língua inferior às línguas orais. O reconhecimento de seu status linguístico, com estrutura linguística e gramatical, se deu apenas em meados do século XX com a divulgação de resultado de pesquisas acadêmicas que demonstravam sua importância para o desenvolvimento cognitivo de surdos. Os estudos de Stokoe, linguista norte americano que na década de 1960 pesquisou a *American Sign Language (ASL)*, representam um marco para esta área de conhecimento. Em 2005, França e Brasil reconhecem suas respectivas línguas de sinais, Língua de Sinais Francesa (LSF) e Língua Brasileira de Sinais (Libras) como as línguas para expressão e comunicação dos surdos dos respectivos países.

No Brasil, a Lei nº 10.436 de 2002, regulamentada pelo Decreto nº 5.626, 2005, estabelece a Libras como língua de comunicação dos surdos brasileiros e como canal de acesso à educação.

Art. 23. As instituições federais de ensino, de educação básica e superior, devem proporcionar aos alunos surdos os serviços de tradutor e intérprete de Libras - Língua Portuguesa em sala de aula e em outros espaços educacionais, bem como equipamentos e tecnologias que viabilizem o acesso à comunicação, à informação e à educação.

Segundo Gesser (2009, p. 77), pesquisadora surda, através da língua de sinais os surdos adquirem e partilham informações que tornam possível a compreensão do mundo. Para Campello (2007) a visualidade se configura como aspecto fundamental na comunicação e, principalmente, na educação dos surdos e as tecnologias têm potencializado o desenvolvimento de materiais pedagógicos adequados à percepção visual, impactando a educação de surdos.

Apesar de seu reconhecimento recente, a Libras tem conquistado espaços importantes em ambientes educacionais de todos os segmentos, em grupos de pesquisa e produção acadêmica. O quantitativo da produção acadêmica e científica é significativo em levantamentos em bases brasileiras como Biblioteca Brasileira Digital de Teses e Dissertações (1.592 teses e dissertações sobre surdos), no Diretório de Grupo de Pesquisa do CNPq (149 grupos que pesquisam sobre surdos) e em repositórios institucionais (1.293 trabalhos no RI da UFSC) e de revistas (293 artigos na SciELO).

A produção acadêmica e de materiais pedagógicos está para além destes espaços e encontra-se presente em todas as regiões brasileiras, dispersa em

diferentes ferramentas. Considerando a necessidade de disponibilizar materiais educacionais em Libras visando a garantir os direitos dispostos na Lei nº 10.436 de 2002, para atender aos diversos segmentos educacionais no Brasil, o Instituto Nacional de Educação de Surdos (INES) desenvolveu um repositório para agregar a produção de diversas instituições com acesso livre para uso e compartilhamento.

Apesar dos avanços e conquistas na educação de surdos, há ainda uma demanda significativa de ferramenta para disponibilizar materiais produzidos para educação de surdos e conhecimentos resultantes de pesquisa sobre a temática. Esta pesquisa tem como objetivo principal discutir a disseminação da informação e conhecimento sobre e para educação de surdos via repositório bilíngue (Português e Libras).

Repositório digital HUET: do textual ao imagético

O projeto se originou da demanda de ampliar o acesso de professores e alunos surdos a “diferentes objetos de aprendizagem para surdos em Libras e Português, agregando materiais desenvolvidos por diversas instituições em um único sistema que permite busca, seleção, uso e reuso” (CHALHUB, 2015). Desta forma foi planejado um repositório temático para educação de surdos.

A elaboração do projeto previu a participação de professores, tradutores e intérpretes (surdos e ouvintes) que tivessem experiência na área da educação de surdos e com alunos do curso de Pedagogia do Departamento de Ensino Superior (DESU) do INES. Foi fundamentado nos depoimentos destes sujeitos e na literatura sobre educação de surdos, materiais pedagógicos para surdos, cultura surda, experiências de outras instituições com cursos e materiais para educação de surdos, que definimos que materiais povoariam o repositório, o layout e as comunidades e coleções.

Foram seguidas as etapas apresentadas por Sales (2011): 1. Seleção de software de gestão de repositórios (*DSpace*) além de aquisição de equipamento, instalação e parametrização/customização; 2. Planejamento do repositório; 3. Definição de políticas; 4. definição de materiais digitais a serem incluídos; 5. Definição das coleções; 6. Definição de serviços oferecidos; 7. Implantação do repositório; e 8. capacitação e divulgação do repositório. Para cumprirmos estas etapas (a última ainda está em andamento), contamos com a parceria de profissionais responsáveis pelos repositórios da FIOCRUZ e UFRJ e com a participação de diversos setores do INES (DEBASI, DESU e DDHCT). A participação de profissionais de diversas formações possibilitou o desenvolvimento de uma ferramenta diferenciada que atendesse uma comunidade com padrão comunicacional diferente da cultura ouvinte. Foram dois anos de muitas conversas e muitas propostas reescritas, redesenhadas até chegarmos à versão final apresentada na Figura 1.

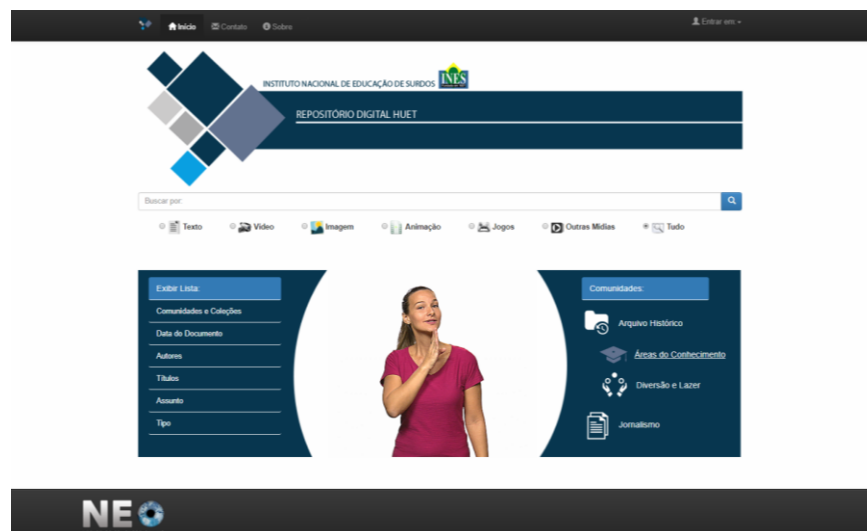


Figura 1: Layout do Repositório Digital Huet com acessibilidade em Libras.

Na Figura 1 é possível perceber o protagonismo do sujeito surdo com a centralidade da comunicação em Libras e da comunicação mais visual e menos textual. Este layout segue a tendência comunicacional de surdos que estão presentes em outros materiais desenvolvidos para esta comunidade linguística, como o curso [Letras Libras](#) da Universidade Federal de Santa Catarina (UFSC).

Nesta página inicial todos os termos em português são apresentados em Libras ao serem tocados pelo mouse. A Figura 1 apresenta a intérprete surda fazendo os sinais em Libras de “Áreas do Conhecimento”, uma das quatro comunidades do repositório. A imagem de abertura do Repositório Digital Huet procura reforçar aspectos visuais visando estimular os surdos (alunos, professores e demais profissionais ligados à educação) a continuar a navegação, seja na busca que pode ser por todos os materiais ou por assunto combinado com o tipo de material, ou ainda na exibição dos materiais via comunidades, autorias, títulos, datas, assuntos e tipos de materiais. Esta redundância de recuperação da informação foi intencional visando potencializar a encontrabilidade dos materiais, que segundo Santos e Madio (2017, p. 282) é importante por “apresentar um menor caminho a percorrer, para que o usuário encontre a informação que procura”.

Como um dos principais aspectos do Repositório Digital Huet é ser uma ferramenta com materiais bilíngues, ele permite a inserção de objetos em diferentes formatos em um mesmo registro e recuperação, conforme Figura 2.

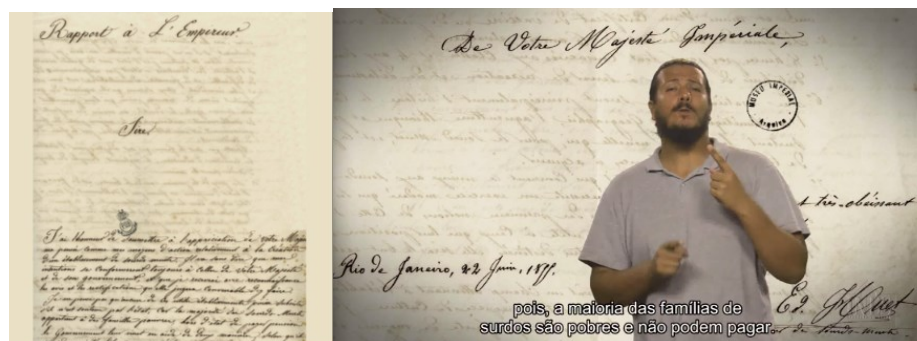


Figura 2: Documento original de 1855 e vídeo com a tradução em Libras

O Repositório foi inaugurado em novembro de 2017 e conta atualmente com mais de 600 objetos, a maioria bilíngue (Libras e português) e teve 4.505.027 acessos até 17 de abril de 2019.

Metodologia

Com o objetivo de avaliar a demanda dos usuários sobre e para educação de surdos via repositório bilíngue (Português e Libras) esta pesquisa descritiva teve abordagem quantitativa e qualitativa. Foi realizado levantamento da estatística de acesso do Repositório Digital Huet visando a identificar os materiais mais acessados. Os dados foram organizados utilizando planilhas Excel e analisados segundo a tipologia, conteúdo, área do conhecimento e instituição responsável pela produção.

Resultados

O Repositório em estudo foi nomeado Huet em homenagem ao professor francês surdo que instituiu a educação pública a surdos no Brasil em 1857. E. Huet chegou ao Brasil na década de 1850 e, com apoio de D. Pedro II, fundou no Rio de Janeiro a primeira escola para surdos com financiamento público atendendo alunos surdos de todo o país, o Instituto Imperial de Educação de Surdos, atualmente Instituto Nacional de Educação de Surdos (INES).

Outras escolas surgiram seguindo os mesmos parâmetros de educação diferenciada para surdos, com materiais específicos, currículos com abordagens no modelo médico, com a oralização dos surdos, ou no cultural, com ensino de língua de sinais, de acordo com as tendências internacionais. Mais recentemente surgiram os cursos superiores com foco na educação de surdos como na Universidade Federal de Santa Catarina, a Universidade Federal do Rio Grande do Sul e no próprio INES.

Os materiais que compõem o Repositório Digital Huet são oriundos de instituições educacionais como as citadas e de museus e arquivos públicos ou privados. A diversidade de materiais que povoa o Repositório em questão reflete

essa diversidade de instituições que produzem materiais sobre educação de surdos e materiais desenvolvidos especialmente para educação de surdos, desde textos até vídeos e imagens (desenhos e fotos).

A Figura 3 apresenta a distribuição por tipo de materiais dos 100 objetos mais acessados, todos com mais de 320 acessos.

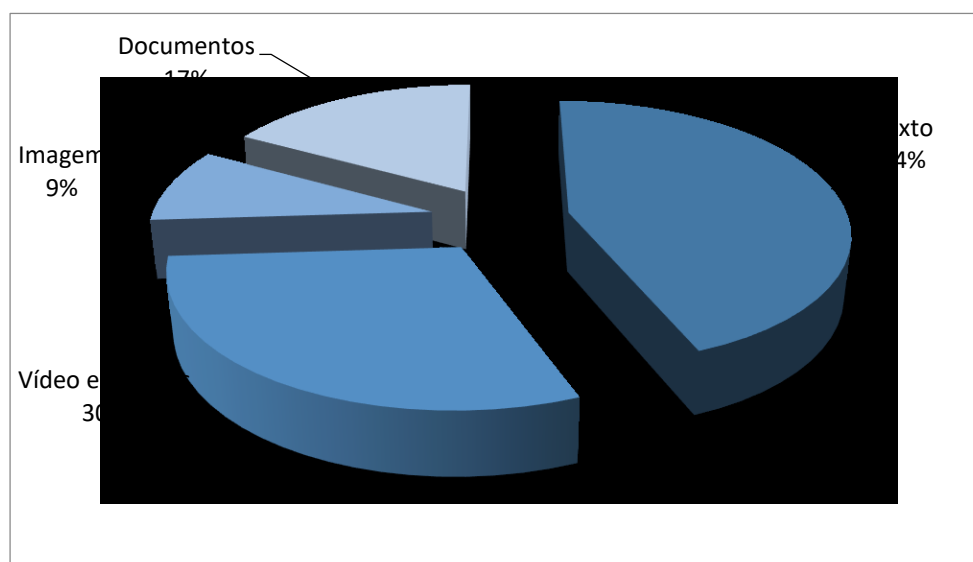


Figura 3: Tipos de materiais dos arquivos mais acessados no Repositório Digital Huet

Entre estes 100 mais acessados, merecem destaque os 10 materiais no topo da lista, sendo nove documentos históricos e uma tese de pesquisadora surda (na décima posição). O documento que é o primeiro em número de acesso é o relatório que Huet enviou a Dom Pedro II, datado em 1855, com as diretrizes de como deveria ser o colégio só para surdos. O [Rapport à l'Empéreur](#) pertence ao Museu Imperial que cedeu uma cópia digital ao INES e o direito de ser traduzido para Libras, sendo disponibilizado em formato de vídeo em Libras e imagem. Outros documentos que estão entre os dez mais buscados são os relatórios do diretor Tobias Leite, carta do Marquês de Abrantes e fotos de alunos em atividades no Instituto.

Considerando os demais materiais, a maioria (44%) é de textos teóricos sobre educação de surdos, seguidos de vídeos (30%). No conjunto de textos, tem destaque os estudos sobre fábulas em Libras, ensino de ciências, a Ata de Milão, a Iconographia (Flausino), acessibilidade em museus e Panorama da Educação de Surdos no Ensino Superior. Entre os vídeos estão presentes uma exposição do Museu Nacional (O Mar Brasileiro na Ponta dos Dedos), literatura infantil (Bela Adormecida e o Curupira), Aula de Libras, uma exposição de Física da Casa da Ciência da UFRJ e programa cultural (Café com Pimenta), programa de documentários esportivos (Super Ação) e reportagens.

Com relação ao tipo do conteúdo, a Figura 4 apresenta um enfoque maior nos arquivos com conteúdo teórico sobre a prática educacional.



Figura 4: Tipo de conteúdo dos arquivos mais acessados no Repositório Digital Huet

De acordo com a Figura 4, há uma demanda grande por materiais teóricos sobre educação de surdos, principalmente artigos e apresentações em eventos científicos. Muitos destes arquivos são pesquisas sobre ensino de língua portuguesa para surdos, a construção de glossários bilíngues em Libras, o ensino de ciências e matemática para surdos e acessibilidade na educação. Alguns destes materiais teóricos estão em formato bilíngue, texto em português com vídeo em Libras.

Os arquivos de conteúdo histórico são fotos e documentos como relatórios, cartas do século XIX e revistas da associação de professores de surdos e revistas do INES. Os materiais das categorias infantil, cultural didático e jornalístico estão em Libras, sendo um deles uma animação da literatura infantil (João e Maria). Os materiais jornalísticos são jornais semanais, reportagens e documentários sobre diversos temas, de tecnologias a esportes. Os materiais didáticos apresentam conteúdo em Libras sobre ensino de Química, Física e Biologia.

Há uma relação direta entre acesso aos materiais e o povoamento segundo as áreas de conhecimento, com exceção da área de História. Na Figura 5 podemos ver essa distribuição por áreas de conhecimento.

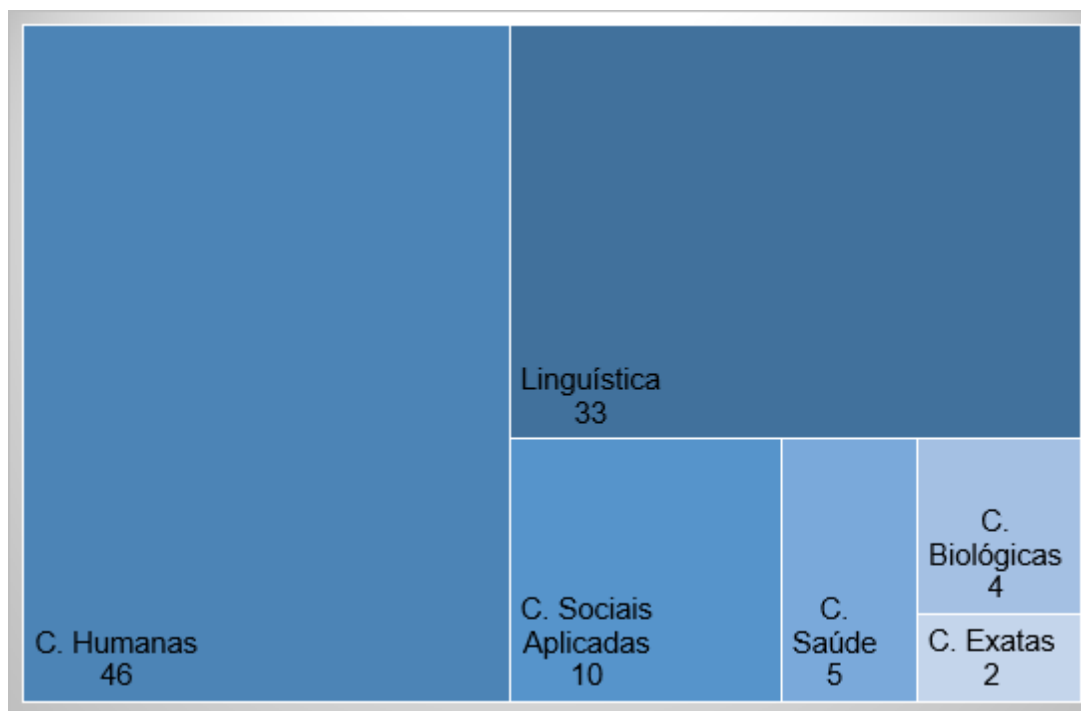


Figura 5: Arquivos acessados no repositório Digital Huet segundo as áreas de conhecimento

As Ciências Humanas é a que apresenta maior número de acessos por área (46) e a segunda em número de arquivos (98) no repositório. Merecem destaque os objetos da História que, apesar de ser uma das áreas com menor número de objetos inseridos (29), teve 24 arquivos acessados, estando nove destes entre os dez de destaque.

A Linguística é a área mais povoada (143) do Repositório e a segunda em número de objetos acessados (33). Vale ressaltar que a maioria dos objetos acessados na área da Linguística está relacionada a textos teóricos sobre ensino de português como L2 para surdos, vídeos de aulas de Libras e vídeos de literatura infantil e lendas brasileiras.

As temáticas presentes nos trabalhos mais acessados são ensino de ciências, pesquisas sobre material didático, currículos, políticas públicas, práticas pedagógicas e letramento.

As áreas com menores números de acesso são Ciências da Saúde, Biológicas e Exatas. Porém, vale ressaltar que alguns destes arquivos apresentam alta frequência de acesso, como os vídeos das exposições de Química e de Física (ambos da casa da Ciência da UFRJ) e o Mar nas Pontas dos Dedos do Museu Nacional (UFRJ).

Considerações finais

As iniciativas de acesso livre à informação são fundamentais para todos os grupos sociais. Todo cidadão deveria ter o direito de acessar informações científicas e educacionais em sua própria língua, mesmo pertencendo a uma minoria linguística, mesmo que sua língua seja visual e não oral.

Os surdos têm o direito de ter acesso ao conhecimento em sua própria língua para se desenvolverem cognitivamente. Desta forma, a construção de um repositório que agregasse resultados de pesquisa, reflexões teórico-prática e material pedagógicos para a educação de surdos se configurava como um projeto relevante e urgente.

Os resultados desta pesquisa demonstram que há uma demanda significativa de objetos de aprendizagem, de materiais históricos e culturais construídos, em sua maioria, na perspectiva da visualidade da comunicação de surdos. O *Repositório Digital Huet* tem possibilitado maior acesso de materiais, disponibilizando em um único espaço objetos de instituições como o Museu Imperial, UFRJ e demais universidades, além do próprio INES. Entre os documentos mais acessados estão os históricos, objetos de aprendizagem, teses, relatórios de pesquisas, vídeos de literatura infantil e de programas culturais. Alguns destes materiais foram produzidos com objetivos educacionais e outros apresentam características que enriquecem o processo de aprendizagem de diversos segmentos educacionais.

Referências

- BASSO, S. P.; CAPELLINI, V. M. L. Material didático para alunos surdos: a literatura infantil em libras. *Revista Eletrônica de Educação*, v. 6, n. 2, nov. 2012.
- BRASIL. *Decreto nº 5.626*, de 22 de Dezembro de 2005. Regulamenta a Lei nº 10.436, de 24 de abril de 2002, que dispõe sobre a Língua Brasileira de Sinais - Libras, e o art. 18 da Lei nº 10.098, de 19 de dezembro de 2000. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2005/decreto/d5626.htm
Acesso em: 21 mar 2019.
- CAMPELLO, A. R. S. *Aspectos da visualidade na educação de surdos*. 2008. 228f. Tese (Doutorado em Educação) – Universidade Federal de Santa Catarina, Florianópolis. 2008.
- CHALHUB, T. *Repositório digital para educação bilíngue Libras-Português*. Rio de Janeiro: INES, 2015.
- GESSER, A. *Libras: que língua é essa? Crenças e preconceitos em torno da língua de sinais e da realidade surda*. São Paulo: Parábola Editorial, 2009.



SACKS, O. *Vendo vozes: uma viagem ao mundo dos surdos*. São Paulo: Companhia da Letra, 2015.

SALES, Luana. *Desenvolvimento de repositório institucional para o Instituto de Engenharia Nuclear*. Rio de Janeiro: Instituto de Engenharia Nuclear. 2011.

SANTOS, J. M. P. dos; MADIO, C. de C. Encontrabilidade da informação: uma análise aplicada no repositório Arca. In: Anais ... COAIC Colóquio em Organização, Acesso e Apropriação da Informação e do Conhecimento. 24 e 25 de agosto de 2017, UEL, Pelotas, 2017. Disponível em: <http://www.uel.br/eventos/cinf/index.php/coaic2017/coaic2017/paper/viewFile/512/353> Acesso em: 02 mar 2019.

Desenvolvimento de Comunidades e Coleções por meio do *Design Thinking*: experiência do repositório do Instituto Nacional do Câncer José Alencar Gomes da Silva

Kátia Oliveira Simões¹, Robson Dias Martins², Camila Belo³, Kesya Cristina Silva de Paula⁴

¹ Instituto Nacional de Câncer José Alencar Gomes da Silva.

Email: katia.simoes@gmail.com

² Instituto Nacional de Câncer José Alencar Gomes da Silva.

Email: rdias100@hotmail.com

³ Instituto Nacional de Câncer José Alencar Gomes da Silva.

Email: camila.ferreira@inca.gov.br

⁴ Instituto Nacional de Câncer José Alencar Gomes da Silva.

Email: kesyacristina.s.p@gmail.com

Resumo

O trabalho apresenta o processo de construção e desenvolvimento das comunidades e coleções do Repositório Institucional (RI) do Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA) realizada por meio de uma metodologia ativa de *Design Thinking*. A proposta inclui uma pesquisa exploratória, a definição da jornada do usuário no RI e um laboratório de co-criação, que foi constituído a partir de uma imersão de dois dias, realizado com indivíduos representativos dos diversos segmentos do Instituto. O grupo de trabalho envolvido no laboratório foi constituído por 15 profissionais e pesquisadores das áreas de Ensino, Pesquisa, Gestão, Prevenção e Assistência. O estudo aborda e descreve a experiência inovadora e dinâmica de construção coletiva das comunidades e coleções e enaltece a importância de imersão dos participantes. Considera a definição da arquitetura e a jornada do usuário no RI numa etapa decisiva para a continuidade do projeto. Para tanto, foi realizada uma prototipagem desenvolvida de forma colegiada. Abarca os obstáculos e desafios transformados em oportunidades e apresenta o fluxo de trabalho realizado. Por fim, oferece como produto final, um conjunto de comunidades representativas que possuem o intuito de facilitar o processo de registro e recuperação das informações de acesso aberto produzidas pelo Instituto, seu corpo técnico e por seus discentes.

Palavras-chave: Repositório Institucional; *Design Thinking*; Tomada de Decisão; Comunidades e Coleções.

Abstract

The present work presents the process of construction and development of the communities and collections of the Institutional Repository (IR) of the National Institute of Cancer José Alencar Gomes da Silva (INCA) carried out through an active methodology of Design Thinking. The proposal includes an exploratory research, the definition of the user's day in RI and a co-creation laboratory, which constitutes a two-day immersion with individuals representing the various segments of the Institute. The working group involved in the laboratory consists of 15 professionals and researchers in the areas of Teaching, Research, Management, Prevention and Assistance. The study addresses and describes the innovative and dynamic experience of collective construction of communities and collections and praises the importance of immersion of the participants. It considers the definition of the user's architecture and journey in RI a decisive stage for the continuity of the project and its prototyping should be developed collegially. It covers the obstacles and challenges transformed into opportunities and presents the workflow realized. Finally, it offers as final product a set of representative communities whose purpose is to facilitate the process of registration and retrieval of open access information produced by the Institute, its technical staff and its students.

Keywords: Institutional Repository; Design Thinking; Decision Making; Communities and Collections.

Introdução

O Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA) busca exercer plenamente o seu papel de órgão normativo do Sistema Único de Saúde (SUS) do Brasil, coordena e planeja ações aplicadas aos serviços de saúde, principalmente, na elaboração de estratégias, diretrizes e protocolos, mais especificamente, relacionadas com a prevenção e o controle do câncer. O Instituto atua de forma integrada com o Ministério da Saúde na construção de diretrizes o que amplia sua abrangência nas áreas de ensino e pesquisa. Ele está alinhado com uma dinâmica informacional com foco na sistematização e divulgação do conhecimento, o INCA caracteriza-se por um ambiente organizacional favorável à livre circulação, produção e disseminação da informação técnico-científica gerada institucionalmente, o instituto adota um modelo de gestão participativa e compartilhada, permitindo a formação de redes de conhecimento técnico-científico para ampliar o conhecimento sobre a doença e desenvolver políticas de saúde pautadas nos princípios de equidade em saúde no contexto do SUS.

O projeto de criação do Repositório Institucional (RI) do INCA, é uma proposta alinhada com o planejamento estratégico da Instituição e envolve um

conjunto de fatores que permeiam seu arcabouço, dentre as quais destacam-se as questões relacionadas com a delimitação das políticas de informação, preservação digital, gestão da informação, usabilidade, suporte tecnológico etc. O recorte deste trabalho relaciona-se ao processo de construção das comunidades e coleções do RI INCA. Atualmente, o debate sobre a disponibilização da informação aberta, com políticas e licenças que garantam essa transparência movimenta as organizações públicas para um envolvimento mais intenso para que esse debate se estabeleça ou se afirme. Sendo assim, o INCA assume seu papel no bojo dessa discussão para assegurar seu compromisso em oferecer sistemas de informação para apoio aos serviços que sustentem as atividades institucionais bem como viabilizar informações e serviços que orientem e direcionem os cidadãos para agilizar seus processos e demandas informacionais.

O projeto RI do INCA já com o software instalado, caminha rumo a elaboração/instituição das comunidades e coleções. As comunidades e coleções são estruturas informacionais que representam a organização do RI.

Estabelecimento do Grupo Gestor

De acordo com, Leite (2009, p. 49) para a construção e implementação de repositórios institucionais, “[...] é importante que seja constituída uma equipe capacitada e comprometida com a realização do projeto. Idealmente, uma equipe multidisciplinar [...]. Dada a importância do projeto, em 2018 foi criada a Portaria INCA instituindo o Grupo de Trabalho do Repositório Institucional do INCA. Trata-se de iniciativa que vem sendo conduzida pela Coordenação de Ensino (COENS), e que objetiva a criação de um Repositório Institucional.

O grupo possui representantes dos seguintes setores do INCA:

- a) Coordenação de Ensino (COENS)
- b) Coordenação de Prevenção e Vigilância (CONPREV)
- c) Coordenação de Pesquisa (COPQ)
- d) Coordenação de Assistência (COAS)
- e) Coordenação de Gestão de Pessoas (COGEP)

Cabe ao Grupo Gestor as seguintes responsabilidades:

- a) definição das políticas e níveis de acesso;
- b) estabelecer critérios para inclusão de conteúdo e povoamento do RI;
- c) identificar possíveis parcerias;
- d) prover orientações sobre o uso do Repositório Institucional e o arquivamento de itens;
- e) definir padrões metadados para a representação dos documentos;
- f) realizar e apoiar as ações para a promoção do Repositório Institucional junto à comunidade do INCA e fora dela e;

- g) Planejar e criar comunidades, subcomunidades e coleções.

Com o Grupo Gestor estabelecido, realizou-se um levantamento preliminar para se elencar quais os critérios para determinar como se constituiria as comunidades, subcomunidades ou coleção no RI. O grupo se deparou com as seguintes questões: Como a criação de comunidades e coleções refletirá a instituição, e como poderá ser base em sua estrutura organizacional? Para responder as questões, foi realizada uma dinâmica com as características de Cocriação, um trabalho de construção conjunta que favorece o diálogo entre os envolvidos para promover o engajamento, inovação, e criação de processos que resultem no estabelecimento do projeto alvo.

Criação de comunidades e coleções para o repositório do INCA

O IBICT através de sua wiki (IBICT, 2013) apresenta a definição de comunidades, subcomunidades, coleções e itens a serem utilizados pelos repositórios brasileiros.

Comunidades e subcomunidades são estruturas informacionais que representam a organização do repositório. As comunidades são as estruturas de mais alto nível e podem conter vários níveis de subcomunidades. Assim, representam apenas a estrutura, não contendo objetos digitais diretamente (IBICT, 2013).

Coleções são os documentos agrupados nas coleções, e as comunidades, por sua vez, agrupam subcomunidades e coleções. Nesse contexto, as comunidades e subcomunidades podem representar temas ou estruturas organizacionais, ou seja, as coleções são estruturas que servem, preferencialmente, para agrupar documentos com alguma característica comum. Toda coleção deve pertencer a uma comunidade ou subcomunidade, pois enquanto as comunidades organizam o repositório, as coleções organizam os documentos do acervo.

Item, por sua vez, é um conjunto de descrições e objetos digitais. Pode-se dizer que é a unidade informacional do DSpace, consiste de vários campos descritivos aliados aos objetos digitais, que unidos formam uma unidade. Os itens são depositados nas coleções, que por sua vez, estão contidas nas comunidades e subcomunidades formando a estrutura do DSpace (IBICT, 2013).

As regras para criação e controle de comunidades, subcomunidades, e coleções variam entre as instituições. Elas podem estar alinhadas com a departamentalização da instituição responsável ou alinhadas com as temáticas do repositório e servem para organização e administração dos RIs. Além disso, colaboram com o seu desenvolvimento e gestão e, delimitam, juntamente com a política do repositório, com o que é permitido ou não ser incluído no RI.

As iniciativas para construção de comunidades e coleções para repositórios institucionais têm permitido, no âmbito profissional para os bibliotecários, novas

experiências de criação de controle e gerenciamento para os recursos eletrônicos, permitindo e possibilitando novas formas de atuação concomitante com diversos profissionais de outras áreas. Em um cenário de inovações e oportunidades e com um quantitativo diverso de ferramentas colaborativas o grupo de trabalho do Repositório Institucional do INCA que optou pela construção de seu repositório em DSpace, considerou pertinente a escolha do *Design Thinking* como ferramenta para a discussão acerca das comunidades e coleções.

Metodologia *Design Thinking*

O mercado contemporâneo apresenta diversas ferramentas que colaboram com os gestores públicos na solução de problemas específicos. No contexto, de solucionarmos a questão relacionada com a construção de comunidades e coleções para o Repositório Institucional do INCA, o grupo de trabalho do RI optou por trabalhar com a ferramenta de *design thinking*. A escolha foi realizada considerando sua abordagem criativa e suas funcionalidades para a solução do problema apresentado, colocando o cidadão como foco principal para a construção das comunidades e, facilitando dessa forma, a recuperação das informações disponibilizadas no repositório.

O termo *design thinking* foi introduzido no início dos anos 2000, por David Kelley, sua ideia foi introduzir o design em estratégias de negócios, tornando-o um processo de transformação como foco em processos organizacionais e em estratégias de pensamento voltado para a criatividade. Na atualidade, essa metodologia busca solucionar problemas, inspirar a criatividade e estimular a inovação com foco no usuário.

Nosso estudo verificou a ausência de estudos relacionando *design thinking* com a criação de comunidades e coleções para repositórios institucionais. Esse cenário ocorre, principalmente, por serem áreas novas e distintas (aparentemente). Em suma, podemos definir *design thinking*, nas palavras de Martin (2009), como um processo de inovação centrada nos usuários, realizado através da colaboração, interação e abordagens práticas com o intuito de solucionar um problema através das melhores ideias e com soluções práticas viáveis. Cardon (2010) complementa afirmando que o *design thinking* é uma “ferramenta útil que aplica o pensamento criativo e crítico para compreender, visualizar e descrever os problemas complexos ou mal estruturados” buscando de forma prática solucioná-los.

A ideia principal da imersão do grupo de trabalho na utilização do *design thinking* foi diagnosticar o problema, explorá-lo, co-criar possibilidades de resolução e implementar a proposta criada pelo grupo ao Repositório.

A metodologia envolveu o trabalho de *Design Thinking*, que parte de um conjunto de ideias e insights para trabalhar o desenvolvimento de proposta futuras, aquisições de informações, análise de conhecimento dentre outras iniciativas. Essa abordagem, combina a capacidade, empatia, criatividade no contexto do problema, inserindo as pessoas no centro da resolução e ou

instituição de um projeto. A dinâmica integra exploração, experimentação e avaliação de ideias, cenários que exploram o conhecimento dos sujeitos envolvidos. A proposta foi realizada em quatro etapas: Imersão; Análise e síntese, Ideação e prototipagem.

Laboratório de cocriação

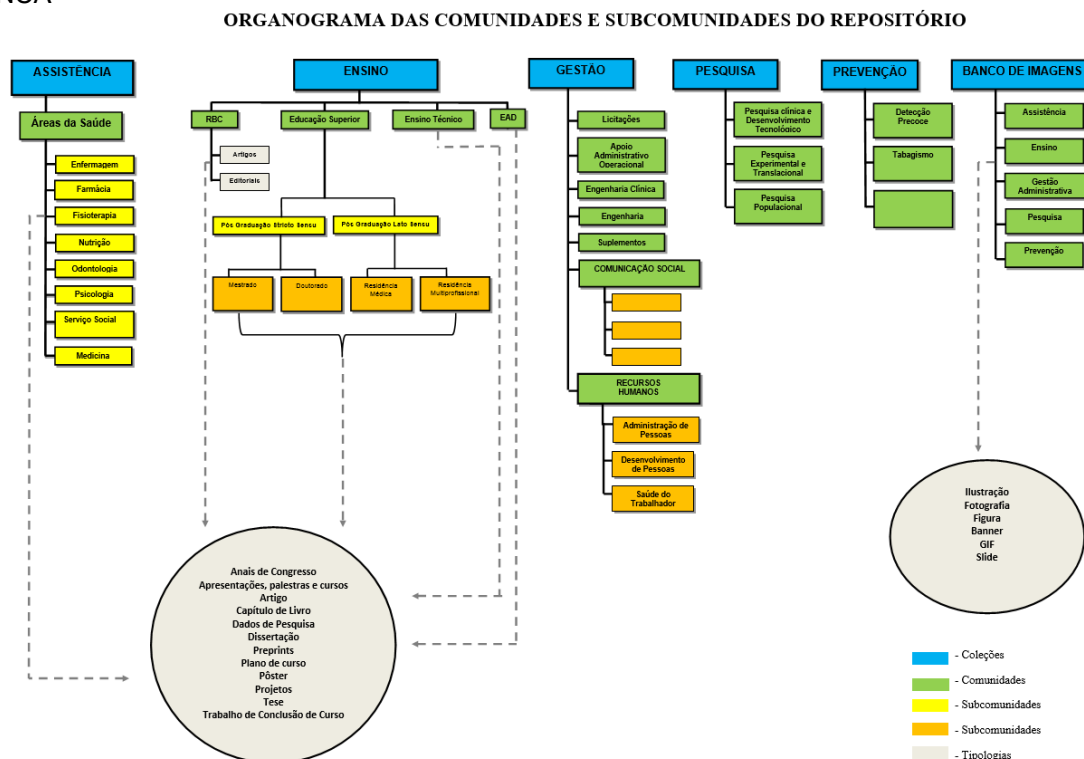
O grupo gestor do RI se reuniu em quatro momentos para o desenvolvimento e estabelecimento das comunidades e coleções. No primeiro encontro foi apresentado o conceito de RI no contexto do acesso aberto, sua importância na comunicação científica, bem como seu papel, objetivos e estrutura no âmbito institucional. Esse momento serviu para o entendimento e alinhamento estratégico entre o grupo, identificação de necessidades e oportunidades que nortearam o trabalho de criação das comunidades. Foi realizada uma análise das áreas institucionais com a perspectiva de traçar o perfil das comunidades e coleções do repositório, e a sua aproximação aos objetivos da organização.

No próximo encontro foi trabalhado o perfil criado com as áreas institucionais idealizados no processo de imersão. Foi detectado mais áreas com afinidades e similaridades. Para facilitar a comunicação dos resultados foi criado um mapa conceitual com os critérios e hierarquia relacionados as áreas elencadas no processo e que reflitam a produção intelectual do INCA. Houve impasse na condução da atividade, pois o INCA abarca diversas frentes de atuação tendo por competência o desenvolvimento de ações nacionais de prevenção e controle do câncer que culminam em diversas pesquisas e, conseqüentemente, várias publicações de diversas áreas. Foi novamente discutida a estrutura de comunidade e coleções e seu papel no processo de armazenar, organizar, disseminar o conhecimento científico, sua idealização deve dar suporte a estrutura institucional. Foi proposto uma análise com base no organograma institucional.

Foram realizados mais dois encontros, onde o grupo idealizou dois protótipos com base no organograma geral do INCA. Foi elencado as funcionalidades do protótipo e como as pessoas e o processo se relacionam.

Por meio dos encontros e análises situacional do organograma se chegou a um primeiro protótipo, no entanto ainda restam algumas lacunas a se observar, para isso estão agendados outros encontros para esta discussão.

Figura 1 – primeiro protótipo de comunidades e coleções para o RI-INCA



Fonte: Grupo Gestor

A utilização da metodologia foi considerada positiva com resultados iniciais satisfatórios. As comunidades e coleções do RI INCA, em seu primeiro protótipo, foi idealizada. O resultado ainda está em discussão com as diversas áreas do INCA, pois esse espelho inicial das estruturas de comunidades e coleções pode ser remodelado conforme a necessidade institucional. Isso ajuda a fomentar a colaboração e a criatividade, agregando diferentes conhecimentos e perspectivas ao projeto RI.

Considerações finais

A partir da análise das estruturas da organização institucional foi possível estabelecer um primeiro protótipo de comunidades e coleções para estruturar o RI-INCA. Essa primeira ideia de organização, até o momento, é aquela que está representando tanto a estrutura organizacional da instituição. O protótipo, no entanto, está aberto a modificações, tendo em vista o processo dinâmico da instituição que atua em várias frentes: assistência, ensino, pesquisa e prevenção.

A combinação das técnicas e instrumentos fornecidos pela metodologia de *Design Thinking* para o desenvolvimento e criação da comunidades e coleções do

RI-INCA foi fator positivo, visto que possibilitaram a identificação e coleta inicial dos dados referentes às estruturas de organização, o esclarecimento de questões que envolvem a criação, modelagem e análise situacional para estruturar um sistema de gestão da informação, como um repositório institucional. O grupo gestor, com uma visão mais ampla, caminham para garantir alcançar, com sucesso, os objetivos traçados para o RI-INCA.

Referências

BROWN, T. **Design Thinking**: uma metodologia poderosa para decretar o fim das velhas ideias. RJ: Elsevier, 2010.

CARDON, E. C. Unleashing design: planning and the art of battle command. **Military Review**, v. 90, n. 2, p. 2-12, 2010.

JULIANI, J. P.; CAVAGLIERI, M.; MACHADO, R. B. Design thinking como ferramenta para geração de inovação: um estudo de caso da Biblioteca Universitária da UDESC. **InCID: R. Ci. Inf. e Doc.**, Ribeirão Preto, v. 6, n. 2, p. 66-83, set. 2015/fev. 2016.

LEITE, F. C. L. Como gerenciar e ampliar a visibilidade da informação científica brasileira: repositórios institucionais de acesso aberto. Brasília: IBICT, 2009.

MARTIN, R. L. **The design of business**: why design Thinking is the Next Competitive Advantage. Boston, MA, USA: Harvard Business Press, 2009.

MELO, Adriana; ABELHEIRA, Ricardo. **Design thinking & thinking design**: metodologia, ferramentas e uma reflexão sobre o tema. São Paulo: Novatec, 2015.

RAMÍREZ Diana Marcela Bernal; ZANINELLIO, Thais Batista. O uso do design thinking como ferramenta no processo de inovação em bibliotecas. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Santa Catarina, v. 22, n.49, p. 59-74, maio/ago., 2017.

Interoperabilidade de repositórios digitais: Em busca da preservação digital

Jean Frederick Brito Xavier¹

¹ Doutorando do Programa de Pós-Graduação em Ciência da Informação da Universidade de São Paulo (USP), Mestre em Ciência da Informação Mestre em Ciência da Informação pela Universidade Federal do Rio de Janeiro / Instituto Brasileiro de Informação em Ciência e Tecnologia UFRJ-IBICT, Arquivista da Universidade Federal de São Paulo (UNIFESP). Email: jeanfrederick3@hotmail.com

Resumo

Introdução: Com o avanço das Tecnologias de Informação e Comunicação, nossa sociedade tem se transformado e esse processo impacta na forma de gerir e disponibilizar as informações. Os Repositórios Digitais têm se destacado por serem ferramentas que auxiliam o acesso e cada vez mais estão sendo relacionados a Preservação Digital. Esse trabalho busca levantar os pontos marcantes sobre esses dois grandes temas e se enquadra no eixo temático “Repositórios Institucionais e Temáticos: Gestão e curadoria, Interoperabilidade, Visibilidade, Preservação digital e Avaliação de repositórios”. **Problema e Justificativa:** Repositórios institucionais e temáticos são parte da solução sistêmica que buscamos para garantir a preservação e o acesso autêntico das informações produzidas por nossa sociedade. Esse problema tem permeado as discussões dos profissionais da Informação, tendo em vista a necessidade de tratar a informação digital. **Objetivo:** Discorrer sobre os aspectos que relacionam os Repositórios Digitais com o campo da Preservação Digital e buscar evidenciar a convergência entre ambos. **Metodologia:** A metodologia da pesquisa exploratória foi empregada neste artigo com o objetivo de proporcionar um panorama amplo sobre o tema. Foi realizada revisão bibliográfica e levantamento de dados sobre Preservação Digital e uso de Repositórios Digitais. **Resultados:** Constata-se que os Repositórios Digitais tendem a incorporar conceitos da Preservação Digital, tal como é disposto no modelo OAIS. **Conclusões:** A solução para a preservação da informação digital não está na aplicação de técnicas isoladas, mas sim no planejamento de procedimentos e estratégias encontradas teoricamente no OAIS e na prática, no Archivematica.

Abstract

Introduction: With the advancement of Information and Communication Technologies, our society has been transformed and this process impacts on how to manage and make available the information. The Digital Repositories have stood out because they are tools that help the access and more and more are being related to Digital Preservation. This work seeks to raise the important points about these two major themes and falls within the thematic axis "Institutional and

Thematic Repositories: Management and Curatorship, Interoperability, Visibility, Digital Preservation and Repository Evaluation". **Problem and Justification:** Institutional and thematic repositories are part of the systemic solution that we seek to guarantee the preservation and authentic access of the information produced by our society. This problem has permeated the discussions of information professionals, in view of the need to treat digital information. **Objective:** To discuss the aspects that relate the Digital Repositories with the field of Digital Preservation and to seek to show the convergence between both. **Methodology:** The exploratory research methodology was used in this article with the aim of providing a broad panorama on the theme. A bibliographic review and data collection on Digital Preservation and use of Digital Repositories was carried out. **Results:** Digital repositories tend to incorporate Digital Preservation concepts, as set out in the OAIS model. **Conclusions:** The solution for the preservation of digital information is not in the application of isolated techniques, but in the planning of procedures and strategies found theoretically in OAIS and in practice in Archivemática.

Introdução

A humanidade tem produzido documentos em escala crescente desde a invenção da imprensa com Gutenberg até os dias atuais. Esse fenômeno ganhou impulso principalmente após o advento dos computadores e dos documentos digitais, que, após a segunda guerra mundial, tiveram seu uso expandido das organizações militares para a população em geral, alcançando as instituições públicas e privadas. A evolução das Tecnologias de Informação e Comunicação (TIC) resultou em uma mudança paradigmática na forma da sociedade gerenciar as suas informações. Gradativamente, com o desenvolver dessa nova dinâmica, os meios tradicionais de comunicação, como o texto impresso, passaram a ser substituídos pela tecnologia digital que proporciona a otimização do processo de produção e divulgação da informação.

A informática traz vantagens como, por exemplo, a facilidade de produção documental em larga escala. Por isso, tornou-se gradativamente mais complexa a tarefa de disponibilizar as informações, seja para o agente da administração, que produz e gerencia as informações, como também, para os pesquisadores, que buscam acessar as informações para fins culturais ou científicos. Sendo assim, após esse primeiro momento de euforia motivado pelos avanços tecnológicos foi constatada a necessidade de se buscar formas para garantir o acesso à massa informacional produzida pelos indivíduos, pelas organizações e governos.

Em resposta às novas demandas para produzir, gerir, disponibilizar e preservar as informações digitais surgem novas soluções como os Repositórios Digitais e softwares de gestão e acesso à informação digital. É importante que as diversas soluções possam interoperar e trabalhar de forma sistêmica e macro, como uma única solução. Com um papel de destaque nesta discussão estão os Repositórios Digitais, que estão evoluindo e expandindo suas funcionalidades, estreitando consideravelmente a sua relação com a Preservação Digital.

Preservação tradicional x preservação digital

Podemos encontrar em Arquivos, Bibliotecas e Museus documentos em suporte de papel, que resistiram às ações do tempo, do manuseio e do ambiente externo por séculos (muitas vezes mesmo sem nenhuma intervenção ou conservação). No entanto, no caso dos documentos digitais há o constante risco de perda de parte da informação ou, até mesmo, de toda informação contida mesmo que pouco tempo após a sua criação.

Com os documentos produzidos em suportes tradicionais como o papel, existe uma preocupação relativa à preservação, que visa a garantir o documento materializado por séculos, com o mínimo de intervenção. Busca-se sempre evitar intervir no documento e protegê-lo dos ataques de pragas e dos efeitos negativos resultantes das reações químicas, que são desencadeadas com a variação da temperatura e umidade do ar (ambiente externo ao documento).

No meio digital, sabemos que instantaneamente pode-se perder completamente todo o material produzido por uma pessoa física ou jurídica, caso não existam sistemas de resgate, *backup*, e diversos outros requisitos relacionados à segurança da informação. Isso significa que a degradação no ambiente digital pode ser instantânea, muito mais rápida do que no caso dos documentos tradicionais.

Na figura seguinte é apontada a diferença de procedimentos referentes à preservação tradicional e à preservação digital:

Figura 1: Ciclos de vida da preservação tradicional e da preservação digital



Fonte: Adaptado de Blue Ribbon Task Force On Sustainable Digital Preservation And Access (2010, p. 25)

De acordo com a Figura 1, podemos perceber que na preservação dos documentos tradicionais, a preservação é somente pensada no momento em que os documentos chegam às instituições responsáveis pela preservação, como Arquivos, Bibliotecas e Museus. Na Preservação digital podemos perceber

que as "ações de preservação" acontecem desde o momento da criação dos documentos até o momento do seu acesso a longo prazo.

Rondinelli explica a diferença da preservação de documentos em suporte de papel para documentos digitais, principalmente o que tange ao seu processo de deterioração, quando nos afirma que:

Quando se trata de documentos de papel, apesar das diversas possibilidades de perda da informação, podemos sentir um maior conforto, pois a não ser que aconteça algum sinistro, estaremos observando o processo de deterioração do papel que dependerá da qualidade do papel, do ambiente em que ele está acondicionado e diversos outros fatores. Quando tratamos de documentos digitais, nos deparamos com o desafio de tratar de documentos que a qualquer momento podem sumir completamente sem deixar grandes rastros, caso não seja implementada uma política de preservação digital dos acervos (RONDINELLI, 2011, p.55)

De acordo com Ferreira (2006), preservação digital é o conjunto de atividades ou processos responsáveis por garantir o acesso continuado a longo prazo à informação. Consiste assim, na capacidade de garantir que a informação digital permaneça acessível e com qualidades de autenticidade suficientes para que possa ser recuperada integralmente no futuro, recorrendo a uma plataforma tecnológica diferente da utilizada no momento da sua criação

Em 2010, Sayão afirmava que as principais estratégias de preservação digital, que estavam sendo praticadas e pesquisadas, tendo como foco garantir a preservação da informação digital, eram a preservação da tecnologia, a emulação, a migração e o encapsulamento. Segundo o autor esses procedimentos são explicados da seguinte forma:

a) Preservação da tecnologia: estratégia baseada na criação de museus tecnológicos que mantêm equipamentos e *software* obsoletos, de forma que os documentos digitais possam ser processados no seu ambiente original. É uma solução de curto prazo;

b) Emulação: estratégia fundamentada na premissa de que o melhor meio de preservar as funcionalidades e a aparência de um objeto informacional digital é preservá-lo junto ao seu *software* original; dessa forma, o objeto pode ser rodado em plataformas atuais por meio de emuladores, que são programas que criam mímicas do comportamento de *hardware* e sistemas operacionais obsoletos em computadores novos. Essa estratégia tem sido foco de muitas pesquisas e controvérsias;

c) Migração – tem como fundamento a migração periódica de um patamar tecnológico em vias de se tornar obsoleto e/ou de se degradar fisicamente para outro mais atualizado e íntegro, incluindo mídias, ambientes de *software*, formatos e computadores; é a estratégia correntemente mais utilizada pelas organizações (SAYÃO, 2007);

d) Encapsulamento – baseia-se na ideia de que os objetos preservados devem ser auto descritos e encapsulados em estruturas físicas ou lógicas com todas as informações necessárias para que seja decifrado e compreendido no futuro.

Além das soluções citadas, ainda são conhecidas outras soluções pontuais como a criação de museus tecnológicos, preservação da cadeia de bits e o refrescamento. Porém, na atualidade é consenso entre os teóricos da Preservação Digital, a necessidade do uso de uma solução sistêmica e que esteja atrelada a uma Política de Preservação Digital, que contemple uma resposta ampla através de uma política de preservação a ser adotada pelas Instituições

Neste contexto, os Repositórios Digitais adquirem destaque pois não só são responsáveis pela guarda da informação digital, mas pela efetiva preservação das informações digitais. Existem então dois grandes pilares que norteiam as atividades dos Repositórios para serem eficientes e confiáveis: a interoperabilidade e a Preservação Digital conforme modelo OAIS.

Um repositório digital deve estar em conformidade com as normas e padrões estabelecidos, de forma a possibilitar níveis de interoperabilidade com outros repositórios digitais e sistemas informatizados. Podem ser citados como exemplos dessas normas e padrões: o “Open Archives Initiative Protocol for Metadata Harvesting – OAI-PMH”, para coleta de registros de metadados em repositórios digitais; o “Metadata Encoding and Transmission Standard – METS”, para a codificação de metadados descritivos, administrativos e estruturais; o “Encoded Archival Description – EAD”, para a codificação de metadados descritivos de documentos arquivísticos; e os “Padrões de Interoperabilidade de Governo Eletrônico – e- PING”, no caso dos órgãos e entidades do governo federal. (BRASIL, 2015, p. 17).

A interoperabilidade como foi dito, se torna importante tendo em vista soluções que devem se integrar em um sistema único como é cada vez mais nítido na área de Arquivos, Bibliotecas, Patrimonial e inclusive de Gestão Administrativa, em que são contidas coleções heterogêneas, com padrões variáveis de metadados, mas que devem ser administrados de forma padronizada.

Sobre a Preservação Digital, que é o centro deste trabalho, destacam-se os requisitos que buscam garantir a preservação dos formatos como é estabelecido na ISO 14721:2012 - Open Archival Information System (OAIS), que é conhecido em português como SAAI (Sistema Aberto de Arquivamento de Informação). É importante destacar que apesar de ter o termo “arquivamento” ou “open archival” este é um modelo direcionado a qualquer Órgão que produza informação digital e não necessariamente a Arquivos.

Repositórios digitais X OAIS

Podemos perceber claramente em (SALES; SAYÃO, 2015) quando nos dizem que há muitos desafios para as áreas da Ciência da Informação, Biblioteconomia e Ciência da Computação em torno dos repositórios digitais de dados de pesquisa, principalmente em termos de prática biblioteconômica e de aplicação de tecnologias. As bibliotecas especializadas – cumprindo sua missão - começam a expandir e renovar suas habilidades e conhecimentos, centradas em documentos, para estabelecer as bases de uma biblioteconomia de dados

capaz de lidar com os estoques crescentes desses ativos informacionais. A seguir destacamos alguns repositórios conhecidos:

- **Fedora:** *software open source* que fornece um repositório de livre acesso por meio de ampla infraestrutura para o armazenamento, gestão e disseminação de objetos digitais, incluindo o relacionamento entre eles (Sayão; Marcondes, 2009);

- **EPrints:** *software open source*, compreendido como uma das mais usadas plataformas para repositórios institucionais. É classificado como o modo mais fácil e rápido de criar repositórios de acesso livre para base de dados científica (Sayão; Marcondes, 2009).

- **DSpace:** *software open source* que fornece um repositório com funções de captura, distribuição e preservação da produção intelectual e científica, dando visibilidade e garantindo acessibilidade no decorrer do tempo (Sayão; Marcondes, 2009).

Porém, esse trabalho busca destacar a *International Organization for Standardization ISO 14721 – Reference model for an open archival information system* (OAIS) que é o resultado do esforço do *Management Council of the Consultative Committee for Space Data Systems* (CCSDS, 2003) na produção de normas que pudessem auxiliar na preservação digital das informações decorrentes dos trabalhos relacionados às missões espaciais.

Apesar de ter sido desenvolvido para missões espaciais, o OAIS tem relevância ampla, servindo como base para discussões relacionadas ao acesso de documentos digitais e à preservação digital. O OAIS descreve os componentes funcionais que devem compor um sistema de informação focado na preservação digital, prevendo o funcionamento das interfaces internas e externas do sistema, possibilitando uma expansão das possibilidades para o uso de metadados de forma que se garanta a preservação digital.

Com o objetivo de ancorar o debate em aspectos práticos, podemos perceber a teoria do OAIS integralmente em um Repositório Digital, o Archivematica, que é um sistema de preservação digital de software livre, gratuito e de código aberto projetado para manter os dados baseados em padrões de preservação digital e o acesso a longo prazo para coleções de objetos digitais.

O aspecto que diferencia o Archivematica de outras soluções conhecidas no mercado e citados anteriormente, diz respeito principalmente ao fato de atender os requisitos relacionados a Preservação Digital constantes no OAIS.

Esta é uma tendência de todos os repositórios, se tornarem o que conhecemos como repositório digital confiável, que são repositórios capazes de manter autênticos os materiais digitais, de preservá-los e prover acesso a eles pelo tempo necessário. (BRASIL, 2015, p. 15).

Para cumprir essa missão, segundo o relatório “Trusted Digital Repositories: attributes and responsibilities” (RLG/OCLC, 2002), os repositórios digitais confiáveis devem:

- Aceitar, em nome de seus depositantes, a responsabilidade pela manutenção dos materiais digitais;

- Dispor de uma estrutura organizacional que apoie não somente a viabilidade de longo prazo dos próprios repositórios, mas também dos materiais digitais sob sua responsabilidade;
- Demonstrar sustentabilidade econômica e transparência administrativa;
- Projetar seus sistemas de acordo com convenções e padrões comumente aceitos, no sentido de assegurar, de forma contínua, a gestão, o acesso e a segurança dos materiais depositados;
- Estabelecer metodologias para avaliação dos sistemas que considerem as expectativas de confiabilidade esperadas pela comunidade;
- Considerar, para desempenhar suas responsabilidades de longo prazo, os depositários e os usuários de forma aberta e explícita;
- Dispor de políticas, práticas e desempenho que possam ser auditáveis e mensuráveis; e
- Observar os seguintes fatores relativos às responsabilidades organizacionais e de curadoria dos repositórios: escopo dos materiais depositados, gerenciamento do ciclo de vida e preservação, atuação junto a uma ampla gama de parceiros, questões legais relacionadas com a propriedade dos materiais armazenados e implicações financeiras.

Podemos concluir que um Repositório Digital não se limita a apenas uma operação mas sim um complexo sistêmico composto de normas, infraestrutura organizacional, *hardware*, *software*, metadados e procedimentos Técnicos, que visam inclusive a preservação digital para que a informação esteja acessível de forma autêntica ao usuário final.

Archivematica

O Archivematica é um software livre desenvolvido sob perspectiva da preservação digital, ao implementar procedimentos que em outra etapa eram isolados, mas agora são automatizados dentro deste software de forma sistêmica, como por exemplo a emulação e a migração dos dados, evitando a obsolescência tecnológica, a incompatibilidade de formato, versões e suportes.

Principalmente em razão da constante mudança de suportes, torna-se inviável que o profissional da Informação esteja sempre monitorando os formatos e atualizando especificamente grupos diferentes. Com o uso do Archivematica, os arquivos são automaticamente convertidos e são tratados em pacotes, tais como estabelecidos no OAIS (Pacote de Informação de Submissão – *Submission Information Package*, Pacote de Informação de Arquivamento – *Archive Information Package* e Pacote de Informação de Disseminação – *Dissemination Information Package*), facilitando assim, a gestão de grandes acervos digitais.

Os microsserviços que são realizados dentro do software, que são responsáveis pelo diferencial do mesmo diante dos outros repositórios. Tendo em vista, que o Archivematica foi baseado no *Metadata Encoding Transmission*

Protocol (METS), o mesmo gera arquivos METS para registro e consultas nos seus diretórios, assim como também por ser baseado no *Preservation Metadata: Implementation Strategies* (PREMIS) preserva os nomes dos objetos digitais nos metadados do PREMIS. O Archivematica também se baseia no padrão Dublin Core e essa flexibilidade e variedade de padrões de metadados, que auxiliaram na arquitetura do software, faz com que seu potencial de interoperabilidade seja elevado.

É importante que se destaque o microserviço de busca por vírus, que o sistema realiza, para garantir a segurança da informação, sendo assim a cada objeto digital que é inserido, o sistema realiza uma varredura para encontrar possíveis arquivos nocivos, não permitindo a contaminação do acervo.

O Archivematica mantém dois formatos de arquivos digitais: uma matriz para preservação e uma derivada para consulta. O repositório usa um sistema de segurança que a muito tempo já era utilizado em órgãos militares e de inteligência, pois possui uma “zona de segurança”, que é inacessível pelo pesquisador que esteja em uma rede. Desta forma, o usuário nunca poderia ter acesso a imagem matriz, apenas tendo acesso a imagem cópia, então mesmo que algum indivíduo tentasse invadir o sistema, no máximo teria acesso a imagem de pesquisa, sem ter qualquer entrada para a matriz, sob a segurança do repositório.

A consulta no Archivematica pode não ser satisfatória, como nos demais repositórios, porém como este repositório é um software livre, pode ser otimizado e por ser baseado nas normas já descritas, pode facilmente interoperar com outras plataformas de acesso como especialmente acontece com o ICA-ATOM ou o ATOM que são soluções criadas pela mesma Organização que criou o Archivematica, a Artefactual System. É importante destacar que este software foi resultado do esforço de diversos colaboradores como a UNESCO, o arquivo da cidade de Vancouver no Canadá, a Biblioteca da Universidade de British Columbia, o Arquivo do Rockefeller Center, dentre outros.

Conclusão

Dispomos diversas opções de Repositórios Digitais para atender nossas Instituições, cada opção com sua especificidade, porém, tendo em vista as questões levantadas inicialmente neste artigo, percebemos que é inevitável que os repositórios se atualizem para tratar com o máximo de segurança as informações sob sua custódia.

O OAIS é apontado como um modelo de referência para que seja feito realmente um arquivamento de informações digitais e não apenas uma *tag* no arquivo digital sinalizando que foi “arquivado”, sendo assim, o modelo de criação de pacotes informacionais, com metadados e todos os procedimentos citados anteriormente é apontado como o ideal na atualidade para conseguir garantir a acessibilidade e autenticidade da informação digital.

Dado a complexidade das questões atuais, a solução para a preservação da informação digital não pode ser limitada a aplicação de técnicas isoladas, mas sim na implantação de procedimentos e estratégias que encontramos teoricamente no OAIS e na prática, no Archivematica.

Referências

BLUE RÍBBON TASK FORCE ON SUSTAINABLE DIGITAL PRESERVATION AND ACCESS *Sustainable economics for a digital planet: ensuring long term access to digital information*. 2010 Disponível em: <http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf> Acesso em 20 mai. 2019.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). Câmara Técnica de documentos eletrônicos. Diretrizes para a presunção de autenticidade de documentos arquivísticos digitais. Rio de Janeiro: Arquivo Nacional, 2012. Disponível em: <http://www.conarq.arquivonacional.gov.br/media/diretrizes_presuncao_autenticidad_e_publicada.pdf>. Acesso em: 28 mai. 2019.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). Câmara Técnica de documentos eletrônicos. Diretrizes para a implementação de repositórios digitais confiáveis de documentos arquivísticos. Rio de Janeiro: Arquivo Nacional, 2014. Disponível em: <http://www.conarq.arquivonacional.gov.br/media/publicacoes/resol_conarq_39_repositorios.pdf>. Acesso em: 13 mai. 2019.

CÉ, G; FLORES, D. Archivemática e ICA-AtoM (AtoM) como Plataformas do Ambiente de Preservação e Acesso de Documentos Arquivísticos. CENDOC, Revista: Revista Comemorativa aos 48 anos do CENDOC—Centro de Documentação da Aeronáutica. Disponível em: <https://www.academia.edu/33214169/Archivemática_e_ICA-AtoM_AtoM_como_Plataformas_do_Ambiente_de_Preserva%C3%A7%C3%A3o_e_Acesso_de_Documentos_Arquiv%C3%ADstico>. Acesso em: 09 abr. 2019.

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEM (CCSDS). Reference model for an open archival information system (OAIS). Blue Book. Washington, Jan. 2002. Disponível em: <<http://public.ccsds.org/publications/archive/650x0b1.pdf>>. Acesso em: 28 mai. 2019.

FERREIRA, Miguel. **Introdução à preservação digital: conceitos, estratégias e actuais consensos**. Portugal. Universidade do Minho. Escola de Engenharia, 2006

RONDINELLI, Rosely Curi. O conceito de documento arquivístico frente à realidade digital: uma revisão necessária – 2011. Tese (Doutorado em Ciência da Informação) – Universidade Federal Fluminense, Programa de Pós-Graduação em Ciência da Informação, Instituto de Arte e Comunicação Social, Instituto Brasileiro em Ciência e Tecnologia, Niterói, 2011.

SAYÃO, Luis Fernando. Uma outra face dos metadados: informações para a gestão da preservação digital **Enc. Bibli. R. Eletr. Bibliotecon. Ci. Inf.**, Florianópolis, v. 15, n. 30, p.1-31, 2010.

SAYÃO Luís Fernando.; SALES L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. *Informação & Informação*, Londrina, v. 21, n. 2, p. 90-115. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939> Acesso em: 19 abr. 2019.

OCLC/RLG - Online Computer Library Center/ Research Library Group. **Preservation metadata for digital objects**: a review of the state of the art. 2001. Disponível em: <<http://www.oclc.org>>

SAYÃO Luís Fernando.; MARCONDES, Carlos Henrique. Software livres para repositórios institucionais: alguns subsídios para a seleção. In: Sayão *et al* (Org.). *Implantação e gestão de repositórios institucionais*: políticas, memória, livre acesso e preservação. Salvador: EDUFBA, 2009.

Preservação digital e Agenda 2030 da IFLA

***Thaíssa Lage Matias da Fonseca¹, Maria Helena Ferreira Xavier da Silva²,
Luana Farias Sales³, Simone da Rocha Weitzel⁴***

1 Bibliotecária da Universidade Federal Fluminense.

Email: thaissamatias@id.uff.br

2 Bibliotecária da Universidade Federal Fluminense.

Email: mh_xavier@id.uff.br

3 Instituto Brasileiro de Ciência e Tecnologia (Ibict).

Email: luanafsales@gmail.com

4 Universidade Federal do Estado do Rio de Janeiro (UniRio).

Email: sweitzel@unirio.br

Resumo

Discute as atuais visões acerca da preservação digital, à luz das diretrizes da agenda 2030 da IFLA. Objetivou-se delinear o estado da arte sobre a preservação digital, seus conceitos e técnicas. Utilizou-se como metodologia pesquisas de cunho bibliográfico, onde foram utilizados textos nacionais e internacionais, de livros e periódicos em meio digital. Apresenta ainda um quadro elaborado a partir da pesquisa, onde são abordadas as principais técnicas e estratégias de preservação no fluxo dos objetos digitais. Concluiu-se que muitas são as ações para a preservação do conhecimento científico, pois a mesma é o alicerce para o desenvolvimento de sociedades democráticas e a continuidade de novos e incansáveis estudos.

Palavra-chave: Preservação digital. Agenda 2030. IFLA. Documento digital.

Abstract

Discusses the current views on digital preservation, in the light of the IFLA 2030 agenda guidelines. The objective was to delineate the state of the art on digital preservation, its concepts and techniques. Bibliographic research was used as a methodology, where national and international texts, books and periodicals were used in digital media. It also presents a framework developed from the research, which addresses the main techniques and strategies for preserving the flow of digital objects. It was concluded that many are the actions for the preservation of scientific knowledge, since it is the foundation for the development of democratic societies and the continuity of new and untiring studies.

Keyword: Digital preservation. Agenda 2030. IFLA. Digital document.

Introdução

A explosão informacional provocou uma revolução nas formas e métodos de geração, armazenamento, processamento e transmissão da informação. As últimas décadas viram um desenvolvimento extremamente rápido na informatização da sociedade. O fenômeno é tão maciço, e mudanças tão influentes no desenvolvimento da nossa civilização, que agora falamos sobre a Sociedade da Informação. O volume e alcance da informação produzida estão se expandindo cada vez mais.

O acesso à informação, mais que um direito amplamente reconhecido pelos principais tratados internacionais de direitos humanos, é, sobretudo um imperativo ético sobre o qual deve se assentar um estado democrático de direito. Não há de se reconhecer a democracia em um ambiente onde o cidadão não é capaz de acessar a informação. É nesse contexto de direito que surge a Agenda 2030 e a essencialidade da biblioteca nesse processo, pois a mesma busca assegurar o acesso público à informação e proteger as liberdades fundamentais, em conformidade com a legislação nacional e os acordos internacionais.

A comunidade acadêmica vem aprimorando meios de comunicação para suplantar as improbabilidades provocadas pela fragilidade tecnológica. Um desses meios de comunicação é o Repositório Digital, que tem como força motriz solucionar: a organização da produção Acadêmica e Institucional, criar mecanismos de busca e acesso e preservar conteúdos dos periódicos eletrônicos.

Agora, as publicações digitais em muitos países complementam, mas não substituiu, em sua totalidade, a publicação tradicional. Esse aumento na informação tradicional e codificada digitalmente desafia as instituições nacionais e, principalmente, as bibliotecas, que ao introduzir novas soluções para gerenciar informações e conhecimento a mesma tem que continuar a cumprir com suas funções que é de registrar e preservar o conhecimento científico para contribuir hoje e sempre com o desenvolvimento e a continuidade de novos e incansáveis estudos.

Materiais e metodologia

A pesquisa se caracteriza como sendo bibliográfica e exploratória, buscando identificar estudos recentes a respeito da importância do contexto da preservação do objeto digital junto às diretrizes da IFLA (Agenda 2030) nas quais o acesso ao conhecimento registrado tem que estar disponível para todos, uma vez que a informação deve atender as necessidades de indivíduos e grupos desempenhando importante papel no desenvolvimento e preservação de uma sociedade democrática.

Na base Scopus e Portal Capes, uma das fontes de pesquisa utilizadas, foram adotadas estratégias de busca a partir das palavras-chave: Preservação Digital, Agenda 2030, Documento Digital.

A partir dos resultados encontrados, foram selecionados 8 artigos, conforme a sua disponibilidade para download e aderência ao objetivo da pesquisa. Realizou-se a leitura dos artigos para análise do seu conteúdo.

Além da pesquisa exploratória, foram consultadas outras fontes de pesquisa, para estruturar o núcleo de análise desse estudo, que é discutido a seguir.

Discussão

É evidente, que hoje em dia, para sobrevivência dos centros de informação e até mesmo, de seus acervos, são necessárias ações de incorporação de tecnologias, para gerenciar/ armazenar e preservar as informações atendendo a demanda emergente que vem desta "sociedade da informação".

O advento das novas tecnologias, como suporte da informação passa por um processo constante de adaptação para atender as necessidades de armazenagem e disseminação da mesma. Uma biblioteca digital é uma biblioteca especial com uma coleção focada de objetos digitais que podem incluir textos, material visual, material de áudio, material de vídeo, armazenados como formatos de mídia eletrônica, juntamente com meios para organizar, armazenar e recuperar os arquivos e mídia contidos na coleção da biblioteca. Bibliotecas digitais podem variar em tamanho e escopo, e pode ser mantido por indivíduos, organizações, ou afiliada com edifícios estabelecidos físicas da biblioteca ou instituições, ou com instituições acadêmicas. O conteúdo digital pode ser armazenado localmente ou acessado remotamente via internet.

Todas as bibliotecas possuem funções básicas que concentra em: coletar, organizar e disseminar os recursos de informação. Tradicionalmente uma biblioteca é um lugar em que os livros, manuscritos, partituras musicais, ou outros materiais literários e artísticos são mantidos para uso e não para venda. Com efeito, é uma instituição orientada para a cobrança e custódia, onde as pessoas podem fazer uso das instalações. Considerando uma biblioteca digital, a mesma é um conjunto de informação digital, com seu armazenamento feito em máquinas de comunicação para reproduzir, transcrever e ampliar os serviços prestados pelas bibliotecas convencionais.

A biblioteca convencional é aquela em que a maioria dos itens do seu acervo é constituída de documentos em papel. Ela existe desde a invenção da escrita. É claro que, antes do advento da imprensa de tipos móveis, em 1440, o seu acervo era formado por outros tipos de materiais (como o tablete de argila, o papiro e o pergaminho). Uma característica da biblioteca convencional é que tanto a coleção como os seus catálogos utilizam o papel como suporte de registro da informação. Todavia, no final do século XIX, houve uma grande revolução na biblioteca, com a introdução do catálogo em fichas e o abandono do catálogo sob a forma de livro. Nas últimas décadas, o computador tem sido utilizado de forma cada vez mais crescente e, desde os anos 1970, muitas bibliotecas implantaram catálogos em linha, passaram a acessar bancos de dados, iniciaram o uso regular do periódico eletrônico e o acesso a textos completos de artigos de periódicos, a verbetes de enciclopédias e a itens de outras fontes de referência. A partir de 1994, por exemplo, com a implantação da World Wide Web (WWW) e do fenomenal crescimento da Internet, as

possibilidades de acessar e recuperar informações aumentaram de forma nunca antes imaginada (CUNHA, 1999 Apud CUNHA, 2008).

Em outras palavras, uma biblioteca digital é um sistema baseado em computador para a aquisição, armazenamento, organização, pesquisa e distribuição de materiais digitais para acesso do usuário final. Não é apenas uma coleção de material em formato eletrônico; ele inclui uma interface do navegador e, talvez, um espaço virtual e da sociedade. Ele requer menos espaço e os dados podem ser disponibilizados através de redes de comunicação para qualquer lugar, facilitando as pesquisas. O digital não é uma entidade única e, como tal, está ligado aos recursos de muitas dessas coleções.

São bastante visíveis às mudanças ocasionadas pela implantação da automação nos sistemas de bibliotecas nas últimas décadas do século XX e na primeira década do século XXI, essas transformações desenvolveram uma economia nos espaços da biblioteca, onde a informação pode ser acessada em formato eletrônico/ digital e manipulada quando e de onde quiser.

Como dito anteriormente, a explosão bibliográfica alavancou as bibliotecas para a procura cada vez maior de itens em formato eletrônicos, porém, as discussões sobre as possibilidades de escolhas – em linha ou CD-ROM, acarretaria na incerteza da durabilidade dessas mídias. Conforme Márdero Arellano (2004, p. 16):

Com o aumento da produção de informação em formato digital, tem sido questionada cada vez mais a importância de se ter garantia a sua disponibilização e preservação por longos períodos de tempo [...]. Os objetos digitais não podem ser deixados em formatos obsoletos para serem transferidos depois de longos períodos de negligência para repositórios digitais.

Dentre outros problemas, para Beargrie & Greenstein (1998 apud MÁRDERO ARELLANO, 2004, p.18) precauções devem ser tomadas para reduzir o perigo da perda de materiais digitais:

- Armazenar em ambientes estáveis e controláveis;
- Implementar ciclos de atualização (refreshment) para cópia em nova mídia;
- Fazer cópias de preservação (assumindo licenças e permissões de copyrights);
- Implementar procedimentos apropriados de manuseio;
- Transferir para uma mídia de armazenamento padrão.

Ainda, para Bullock (1999 apud MÁRDERO ARELLANO, 2008, p. 46) o IPR (Intellectual Property Rights) é um dos principais entraves na preservação dos objetos digitais. Os direitos de propriedade intelectual (IPR) é uma questão importante, pois tem um impacto significativo sobre a preservação digital. Sabemos que a lei de IPR foi originada e criada há anos, quando não havia pensado a nível mundial Web. Temos ciência, também, que o IPR foi estabelecido para os acervos tradicionais (papel), enquanto que para materiais eletrônicos deixa a desejar. A propriedade intelectual, no âmbito dos materiais digitais, são mais complexos e importantes do que a mídia tradicional e se não forem tomadas decisões como a legalização da responsabilidade para a

preservação digital por parte da biblioteca dos materiais que são adquiridos por ela, teremos um caos informacional. Segundo Muir (2004 apud MOGHADDAM, 2008, p. 85):

[...]libraries other than legal deposit libraries will probably want to take responsibility for digital preservation of material they create or purchase. However, there is a lack of awareness of what the law allows. The current legal situation in the publishing industry needs to be clarified and changes to the law could be considered if necessary. Changing copyright law to facilitate legal deposit is one possibility and this is being pursued in some countries such as the UK, but this will not help the vast majority of libraries because legal deposit collections are collections of last resort.

Não obstante, o conteúdo dos recursos digitais e sua necessidade de software associado devem ser levado em consideração. Conforme Márdero Arellano (2008, p. 47):

Salvar os bits de um objeto de digital é necessário, mas não o suficiente para preservá-lo. É necessário também conhecer os atributos da aplicação na qual ele foi criado e com a qual ele pode ser interpretado. O esforço da preservação engloba também o software e o hardware no qual o objeto digital pode ser executado, já que eles também podem ficar obsoletos.

Vale considerar a necessidade de haver equipamentos de leitura para essas versões, e, também treinamento tanto para os bibliotecários como para os usuários, já que não é fácil manter-se atualizado com as mudanças que ocorrem.

Assim, para resguardar a informação digital produzida da destruição causada pelo tempo e pela obsolescência da tecnologia, é necessário método apropriado para tal finalidade. Estratégia de preservação requer “(...) o planejamento, alocação de recursos e aplicação de métodos e tecnologias para assegurar que a informação digital de valor contínuo permaneça acessível e utilizável” (HEDSTROM, 1996 apud MÁRDERO ARELLANO, 2004, p.17) como demonstrado no quadro a seguir:

Quadro 1: As estratégias de preservação no fluxo dos objetos digitais na biblioteca digital

Aquisição	Entrega pelo produtor	
	Captura pela biblioteca	
	Coletado pela biblioteca	
Verificação)	Integridade física (meio)	
	Integridade do conteúdo	Integridade lógica
		Autenticação
Registro	Metadados	Descrição bibliográfica
		Instalação e manipulação
		Acesso
		Preservação

Preservação	Preservação física	Refrescamento do meio
	Preservação lógica	Migração de suporte
	Preservação intelectual	Conversão de formatos Emulação
Acesso	Condições de uso	Acesso local
		Acesso remoto

Fonte: (BORBINHA & CORREIA, 2001 Apud MÁRDERO ARELLANO, 2004, p.17).

Portanto, uma vez que a compreensão das informações está diretamente relacionada às suas propriedades, seus suportes e ao aparato de preservação digital, cada um dos itens relacionados nesse estudo deverá ser avaliado de forma peculiar no caso das publicações eletrônicas. Os suportes que os caracterizam têm pontos em comum, mas também outros bastante distintos.

Muitas bibliotecas estão assumindo esse compromisso. Esse aumento exponencial da informação não se limita à publicação; Aplica-se ainda mais aos dados nos setores acadêmicos e de pesquisa, particularmente nas ciências. As publicações são agora um aspecto da cultura popular e do recorde cultural. Conforme afirma Borba (2009, p.17), a preservação digital tem a ver com a conservação e preservação do patrimônio cultural da humanidade, cuja atenção antes estava voltada apenas para os registros em suportes físicos orgânicos, e que agora se aplica aos formatos de expressão digital.

A informação é base para todas as áreas do conhecimento e cada uma dessas áreas a conceitua de acordo com a sua especificidade. Na Ciência da Informação esta é entendida como um conhecimento inscrito (registrado) em forma escrita (impressa ou digital), oral ou audiovisual em um suporte (LE COADIC, 2004, p. 4). Assim, podemos dizer que a informação tem responsabilidade social na transferência do conhecimento, incluindo aqui o acesso e uso da informação pela comunidade, como forma de suprir suas necessidades e demandas informacionais (WERSIG, NEVELING, 1975).

Conforme a as diretrizes da agenda 2030 da IFLA (2015) as bibliotecas contribuem para o desenvolvimento nacional por meio do acesso a informação e isso quer dizer que a informação é a questão *sine qua non* para o desenvolvimento econômico, político e social de um país. Na sociedade atual a informação é o recurso que mobiliza a economia mundial, sendo o principal elemento-chave na produção das sociedades desenvolvidas.

Para a IFLA (2015) “o acesso público a informação suporta a criação de sociedades de conhecimento que incluem a infraestrutura, as novas tecnologias da informação e comunicação e a capacitação no âmbito da literacia dos media e informativos que o público necessita”. Isso quer dizer que a informação é uma prática social, pois envolve ações de atribuição e comunicação de sentido que por sua vez provocam transformações nas estruturas uma vez que geram novos conhecimentos. Nesse sentido, a

informação é o elemento que torna possível a transição e a transformação do homem na sociedade.

Resultados

A preservação digital requer procedimentos específicos e técnicas apropriadas para cada tipo de formato e mídia, bastante sensíveis e passíveis a qualquer mudança, e como podemos verificar a informação abrange todos os aspectos da vida em sociedade e desempenha um papel importante no desenvolvimento sustentável dessa. Sendo assim, conforme corrobora a IFLA (2015) “o acesso público a informação permite que as pessoas tomem decisões informadas que possam melhorar as suas vidas”, isso quer dizer, que a informação além de ter um poder transformador, ela exerce uma função social na construção do sujeito.

Conclusões finais

Buscou-se por meio desse trabalho realizar uma série de paralelos com a literatura estudada junto com as iniciativas e interações que estão ocorrendo na atualidade. Percebemos que a reflexão a respeito da preservação digital ganhou respaldo e reconhecimento de instâncias maiores, como a IFLA. Diante, de tudo que foi exposto é evidente a importância da preservação digital, pois de acordo com a IFLA (2015, p.13):

Os serviços das bibliotecas contribuem para melhorar os resultados em todos os Objetivos de Desenvolvimento Sustentável (ODS):

- Promovem a alfabetização universal, incluindo a literacia dos media e informação e competências de literacia digital;
- Encerram lacunas no acesso à informação e auxiliam o governo, a sociedade civil e as empresas, a compreender melhor as necessidades locais de informação;
- Fornecem uma rede de sites para programas e serviços governamentais;
- Promovem a inclusão digital através do acesso às TIC com o auxílio de técnicos dedicados que ajudam as pessoas a desenvolverem novas competências digitais;
- Servem como centro de investigação para a comunidade académica;
- Preservam e dão acesso à cultura e património mundial.

Sendo, assim, reflexões como estas objetivam provocar indagações, reações, e por que não mudanças nos sistemas de comunicação científica para que os mesmos sejam mais acessíveis e desta forma, possam cumprir com suas funções que é de registrar e preservar o conhecimento científico para contribuir hoje e sempre com o desenvolvimento da sociedade e a continuidade de novos e incansáveis estudos.

Referências

BORBA, Vildeane da Rocha. **Modelo orientador para construção de estratégias de Preservação digital**: Estudo de Caso do Banco de Teses e



Dissertações da UFPE. 2009. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal da Paraíba, João Pessoa, 2009.

CUNHA, Murilo Bastos da. Das bibliotecas convencionais às digitais: diferenças e convergências. **Perspect. ciênc. inf.**, Belo Horizonte , v. 13, n. 1, p. 2-17, Apr. 2008.

IFLA. **Conjunto de ferramentas**: as bibliotecas e a implementação da agenda 2030 da ONU. Holanda, 2015.

LE COADIC, Yves-Francois. **A ciência da informação**. 2.ed. rev. e atual. Brasília: Briquet de Lemos, 2004. 124 p.

MÁRDERO ARELLANO, Miguel A. **Critérios para a preservação digital da informação científica**. 2008. 356 f. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 2008.

_____. Preservação de documentos digitais. **Ci. Inf.**, Brasília, v. 33, n. 2, aug. 2004 .
MOGHADDAM, Golnessa Galyani. Preserving scientific electronic journals: a study of archiving initiatives. **The Electronic Library**, v. 26, n. 1, p. 83-96, 2008.

SCHWARTZMAN, S. Universidade e pesquisa científica: um casamento indissolúvel? In: SCHWARTZMAN, S., CASTRO, C. M. (Org.). **Pesquisa universitária em questão**. Campinas: Ícone, 1986. p. 11-18.

SILVA, Maria Helena Ferreira Xavier da. **Preservação digital dos periódicos eletrônicos por meio de repositórios digitais**: anteprojeto de pesquisa apresentado ao Programa de Pós-Graduação em biblioteconomia da PPGB. Rio de Janeiro: UniRio, 2017.

WERSIG, Gernot; NEVELING, Ulrich. The phenomena of interesting to information science. **Information Scientist**, v.9, n.4, p. 127-140, Dec. 1975.



Sobre o projeto Biblioteca do IPPUR 2.0

Cláudia Regina dos Anjos¹

¹ Bibliotecária do Instituto de Pesquisa e Planejamento Urbano e Regional (IPPUR) e da Universidade Federal do Rio de Janeiro (UFRJ). Email: cranjos@gmail.com

Resumo

As mídias sociais se tornaram regular em nossas vidas. Vozes influentes sejam para promover um produto, um serviço, ou apenas para manter-se atualizado com as últimas notícias. Com esses elementos digitais um bom número de oportunidades surgem para as Bibliotecas Universitárias (BUs) e novos caminhos se abrem para a comunicação entre bibliotecas e usuários/clientes. Com o uso adequado das redes sociais digitais, as BUs podem atrair e estreitar o relacionamento com os diferentes públicos acadêmicos. Assim, com a consciência de que toda biblioteca tem o dever de atender a 2 (dois) tipos de usuários: presenciais e remotos, a Biblioteca do Instituto de Pesquisa e Planejamento Urbano e Regional (BIBPPUR) criou o Projeto Biblioteca do IPPUR 2.0 - um serviço de referência digital que tem seus produtos e serviços direcionados para informação e educação de usuários/clientes. O projeto foi idealizado para proporcionar maior praticidade na interação e apoio aos usuários/clientes atendendo à 4ª Lei de Ranganathan: "*Poupe o tempo do leitor*". Deste modo, a BIBPPUR através do Setor de Referência e Circulação instituiu redes sociais digitais. No entanto, é importante justificar a implantação e uso das tecnologias 2.0. Para isso deve-se estar alerta às métricas e ao monitoramento. Este trabalho busca nortear a criação, manutenção e monitoramento a partir de uma revisão de literatura sobre o assunto e em conjunto com experiências consolidadas no uso das mídias sociais *Online*.

Palavras-chave: Mídias sociais. *Facebook*. *Twitter*. *Youtube*. Bibliotecas Universitárias. *Web 2.0*. Métricas e monitoramento de mídias sociais.

Introdução: mídias sociais X bibliotecas

As mídias sociais mudaram a vida das pessoas no século XXI - *mídias sociais* é um termo genérico que abrange um conjunto de ferramentas da Web 2.0 e sites de redes sociais como: *Myspace*, *LinkedIn*, *Twitter*, *Facebook*, *Youtube*, *Flickr*, *WordPress*, *SlideShare*, *Google Drive*, *Wikipédia*, *Blogger*, *Skype* e outros serviços que tenham como objetivo o compartilhamento de conteúdo (CALIL JUNIOR, CORRÊA; SPUDEIT 2013).

As ferramentas da *Web 2.0* abriram novas possibilidades de comunicação, agora podemos nos comunicar em muitas novas formas: através de *chats* (bate-papo ao vivo), *links*, fotos, vídeos, etc. Com essas ferramentas as bibliotecas, podem chegar a milhares de usuários, de diferentes tipos, independentemente da localização geográfica. Para Barros (2009), os novos serviços em redes sociais causaram uma

série de transformações na *Web*. A transformação, entretanto, não é baseada apenas em mudanças tecnológicas, mas principalmente, em uma mudança de mentalidade.

Essas tecnologias tornaram mais barato e mais fácil criar material e torná-lo amplamente disponível para um público global, como por exemplo: produzir um livro, um filme, uma música ou uma imagem. O uso dessas mídias modificou a forma como as pessoas se relacionam em sociedade e estão contribuindo para alterar também a maneira tradicional de se comunicar no ambiente biblioteconômico. Na forma tradicional o contato entre a biblioteca e o usuário ocorria só fisicamente, ou seja, na biblioteca, ou por escrito (carta) e depois e-mails.

O objetivo de usar as mídias sociais, em bibliotecas, é particular e pode variar de biblioteca para biblioteca, mas estas tecnologias proporcionam uma alternativa de comunicação direta, informal que pode complementar os outros canais de atendimento ao usuário/cliente, além de ser um sopro de ar fresco na rotina da biblioteca - tanto para os funcionários quanto para os usuários. Os principais argumentos em defesa ao uso desses meios de comunicação (via *Web*) em bibliotecas são:

- fazem parte do dia a dia dos usuários/clientes;
- facilitam a interação ou o contato direto com os usuários;
- dão oportunidade de alcançar usuário e não usuários (da biblioteca), ou seja, elas permitem chegar tanto aos usuários reais quanto aos usuários potenciais.

Desta forma, estes recursos sociais abrem oportunidade para promover as atividades da biblioteca no próprio ambiente dos usuários. Tendo potencial para se tornar a maneira mais eficaz de comunicação e/ou interação da biblioteca com o usuário. Ampliando o horário de atendimento, inclusive para fora do horário de atendimento da biblioteca, estendendo este horário para as 24 horas do dia.

Justificativa: Necessidade de adequação das bibliotecas universitárias frente aos avanços tecnológicos

“Em tempos de mudança, os aprendizes herdarão a terra, enquanto aqueles apegados às suas velhas certezas se descobrirão perfeitamente equipados para lidar com um mundo que já não existe mais” (Éric Hoffer).

O avanço das Tecnologias de Comunicação e Informação (TIC) tem redimensionado cenários em todos os âmbitos da sociedade atual e as bibliotecas acadêmicas não ficaram isentas desse processo, assim, têm-se observado grandes mudanças no âmbito das Bibliotecas Universitárias (BUs) nos últimos anos. Há não muito tempo, os usuários viam o espaço físico da biblioteca como um importante local de pesquisa, mas com a ascensão dos conteúdos digitais, o físico foi transformado em uma rede de opções, com diferentes níveis de acessibilidade e esta transformação mudou a natureza dos trabalhos biblioteconômicos.

Segundo Nascimento (2000) a definição de BUs como sendo um local onde se conservam grandes quantidades de espécies documentais deixou de ser sinônimo de qualidade, pois hoje as coleções informatizadas não são suficientes para garantir a necessidade das BUs e nem também para continuar auxiliando usuários no apoio à pesquisa e nem nas novas formas de comunicação científica. Cunha (2010) observa que as BUs estão deixando de ser a principal fonte de busca de informações dentro

do mundo acadêmico, porque o impacto da *Internet* fez com que as bibliotecas perdessem a supremacia na realização deste papel fundamental.

Behr, Moro e Estabel (2010) sinalizam que o momento é de transição, tanto para a biblioteca quanto para o profissional da informação, uma vez que o ambiente tradicional das bibliotecas medievais cede espaço para bibliotecas que são avaliadas pelos serviços que prestam, pelos produtos que oferecem e pelo uso que fazem das TICs. Consciente do poder que trazem as TICs, as universidades estão começando a se mostrar presentes nos novos canais interativos - como as mídias sociais - para se comunicar com a comunidade universitária.

Assim, hoje é possível observar nas mídias sociais de universidades informações sobre: publicações institucionais, processos de admissão dos cursos, programas de disciplinas, informações de eventos ou palestras, divulgação de fotos e de vídeos de acadêmicos etc.

Diante dos efeitos sociais provocados pelas TICs, Carlos Levi – ex-Reitor da Universidade Federal do Rio de Janeiro (UFRJ) – declarou que a universidade deve se repensar, se reinventar e se reconfigurar para acompanhar as aceleradas transformações, descobertas e conquistas do mundo atual e aproximar as atividades universitárias dos anseios e expectativas da sociedade (A PESQUISA... 2014). Fica claro, nas palavras do reitor, que com o domínio das novas tecnologias, o universo acadêmico está sendo modificado.

Cabe destacar, neste momento, as palavras de Prado (1992), quando esta afirma que a biblioteca acadêmica é uma universidade em si mesma. Então, seguindo essa linha de raciocínio, se as universidades estão modificando suas posturas para acompanhar as transformações do mundo atual, consequentemente as BUs também devem seguir o mesmo destino: se repensar, se reinventar e se reconfigurar já que os objetivos destas são determinados pela própria universidade.

A autora lembra ainda que as BUs têm papel imprescindível na preservação e na disseminação de informações e conhecimento, pois além de atuarem como espaço de inter-relacionamento da comunidade acadêmica em todos os níveis – dos alunos, professores, técnicos e dirigentes das diversas unidades – também estabelecem relações que vão do apoio ao ensino à disseminação da produção científica.

Roly Keating – diretor da Biblioteca Nacional do Reino Unido – em recente entrevista, sobre a influência da ascensão digital no mundo das bibliotecas, disse que o mundo do conhecimento está se tornando cada vez mais digital e baseado em tela, mas isso não implica a diminuição das bibliotecas físicas. O entrevistado lembra que por décadas uma série de valores foram incutidos nas bibliotecas e hoje novos valores estão sendo trazidos na vida dessas instituições por conta dos celulares, dos *iPhones*, dos *tablets*, dos computadores e das redes sociais, e agora é papel da biblioteca para garantirem sua durabilidade unir-se a essas coisas e fazer parte de tudo isso. (KEATING, 2015).

Fernando Arbache - presidente da Arbache (2015) - alerta que modificações estruturais são necessárias para o progresso entre empresa e profissional, segundo ele, cabe aos diretores de planejamento, estratégia e recursos humanos preparar suas organizações para receber a nova geração que dominará o mercado de trabalho em poucos anos: "*As instituições muito tradicionais precisam se adequar e agir de forma mais interativa [...]*" (FRIZON, 2015, p. [4]). É verdade que serviços tradicionais nem

sempre são transferíveis para o contexto digital, mas é inegável que as atividades de uma biblioteca acadêmica estão cada vez mais baseadas em serviços.

Kotler (2000) afirma que empresas e organizações precisam correr mais rápido para continuar no mesmo lugar e que estas devem trabalhar para construir uma imagem sólida e positiva na mente do seu público, enxergando mudanças como oportunidades para que suas estruturas, sistemas e cultura organizacional não corram o risco de ficarem obsoletas. Jesus e Cunha (2012) usam as palavras de Ranganathan para recordar que as bibliotecas são organismos vivos em crescimento, e que existem para atender o usuário e lembram que estas instituições precisam estar atentas para que seus produtos e serviços satisfaçam às necessidades de seus usuários e não se tornem obsoletos. Neste contexto, segundo Viana (2013), o uso das mídias digitais surge como uma alternativa para:

- ajudar pesquisadores a tornar seu trabalho mais visível na *Web*;
- promover o uso autônomo dos serviços;
- promover o acesso autônomo à informação;
- promover o acesso instantâneo à informação;
- promover o aprendizado assíncrono (a aprendizagem assíncrona ocorre em cursos onde o próprio aluno determina a duração, geralmente por intermédio de *CD-ROM*, *Internet*, *e-mail* ou grupos de discussões *Online*);
- promover o aprendizado a distância;
- promover a produção rápida de conhecimento;
- promover o acesso aberto (*e-science*, ciência colaborativa, repositórios institucionais);
- promover o uso de novos dispositivos e de novos aplicativos.

Conscientes deste fato, alguns bibliotecários estão usando as mídias sociais para: ouvirem o que é dito sobre a instituição, para desenvolverem a autonomia de seus usuários e para disseminarem (ou disponibilizarem) acervos dentro e fora de sua biblioteca, entre outras aplicações.

A biblioteca do IPPUR no universo 2.0: o Projeto Biblioteca do IPPUR

Em bibliotecas acadêmicas as atribuições do bibliotecário vão além de sua atuação como agente facilitador da informação. Nesse cenário esse profissional torna-se um profissional de educação, pois em instituições acadêmicas este profissional cria programas de treinamento como: ensino de normalização de trabalhos acadêmicos conforme normas da Associação Brasileira de Normas Técnicas (ABNT) - treinamentos de bases de dados, além dos habituais treinamentos de buscas no catálogo da biblioteca.

Com o mercado cada vez mais concorrido, oferecer melhores produtos e serviços já não é o suficiente, isto porque não há produto ou serviço bom o bastante. Uma maneira de competir no mercado é garantir a satisfação dos clientes. Sendo assim, ouvir as opiniões dos usuários para melhorar o desempenho e fidelizar os clientes são itens essenciais para sobrevivência das instituições e/ou organizações (FIDELIZAÇÃO, 2015).

Antes da *Internet* as bibliotecas, nem sempre tinham essa atitude, talvez porque não fosse tão necessária uma vez que os recursos disponíveis, nestas organizações, eram únicos e não havia outro lugar para encontrá-los. Mas a *internet* mudou tudo

isso e agora as bibliotecas competem também com um ambiente de fontes de informação diversificada e abundante disponível na *Web* e para defender sua própria relevância precisam operar também neste ambiente concorrente.

Recentemente o Instituto Pró-Livro, revelou em uma pesquisa do que: 66% dos brasileiros não frequentam bibliotecas, e outros 14% visitam raramente o local, segundo Dalgiza Oliveira - professora de biblioteconomia da Universidade Federal de Minas Gerais (UFMG) - uma das explicações para a impopularidade das bibliotecas é a falta de visibilidade, onde a própria comunidade do entorno, dessas organizações, não conhece as possibilidades que o espaço oferece (QUESADA, 2016).

Histórico

A Biblioteca do Instituto de Pesquisa e Planejamento Urbano e Regional da Universidade Federal do Rio de Janeiro (UFRJ) - Biblioteca do IPPUR/UFRJ - foi criada em março de 1986, contando inicialmente com recursos da Fundação José Bonifácio (FUJB) e da Financiadora de Estudos e Projetos (FINEP). Em julho de 1988, o acervo, que se encontrava na Biblioteca Central do Centro de Tecnologia, foi transferido para a nova sede do Instituto, no Prédio da Reitoria. Ao longo dos anos, o acervo da biblioteca foi enriquecido através de compras, permutas e doações.

Dentre as doações de acervos de grande relevância, merecem destaque: parte da coleção da extinta Biblioteca do Banco Nacional de Habitação; parte da coleção particular dos professores Michael David Vetter e do Engenheiro José de Oliveira Reis e os acervos particulares dos Professores Eduardo Guimarães de Carvalho e Martim Oscar Smolka. Especializada em Planejamento Urbano e Regional, a Biblioteca busca assegurar suporte ao caráter multidisciplinar de ensino e pesquisa nessa área, através de coleções em vários campos disciplinares, tais como: Economia Urbana e Regional, Geografia, Sociologia, Antropologia, Ciências Políticas, Arquitetura e Urbanismo, História Urbana, História Geral e do Brasil, Meio Ambiente e Filosofia. Hoje entre livros, folhetos, obras de referência, teses e dissertações, vídeos e *Cd-Rom* a biblioteca conta com mais de 24.000 itens.

Projeto Biblioteca do IPPUR

O Projeto Biblioteca do IPPUR 2.0 nasceu da percepção que as mídias sociais poderiam alavancar a performance da biblioteca junto a seus usuários/clientes. Desta maneira, os bibliotecários ocuparam-se primeiramente em repensar sobre serviços e produtos oferecidos e em seguida desenvolver ações pautadas no uso das mídias sociais. Tendo como meta principal a 4ª Lei de Ranganathan "*Poupe o tempo do leitor*" e assim paulatinamente os bibliotecários foram modificando a rotina dos serviços prestados pela biblioteca e passando a produzir e oferecer os produtos e serviços na esfera 2.0.

É necessário frisar que as ferramentas sociais são gratuitas, ou seja, que todo processo de criação, manutenção e hospedagem do site e/ou *Blogs* são inteiramente gratuitos, já que não geram custos para criação e nem para a manutenção dos mesmos, além de não requerem nenhuma habilidade de *Web design* ou de publicação especial.

As ferramentas 2.0 mais conhecidas e usadas para criação de site e/ou *Blogs* são: a *Blogger* e a *WordPress*. Depois de conhecer as duas ferramentas optou-se por usar a *Blogger* - também chamado de *Blogspot* é uma das maiores plataformas de criação de *Blogs* grátis que existe é um serviço do Google, que oferece ferramentas para edição e gerenciamento de *Blogs*, semelhantemente ao *WordPress*, é mais indicado para usuários que nunca tenham criado um *Blog*, ou que não tenham muito familiaridade com a tecnologia.

É verdade que serviços tradicionais nem sempre são transferíveis para o contexto digital, mas a biblioteca conseguiu vencer essa barreira e transformou alguns serviços e/ou produtos presenciais em *Online*. Assim a Biblioteca do IPPUR começou a se lançar no cenário 2.0. A seguir os produtos e/ou serviços criados listados por ordem cronológica de criação:

Blog Auxílio à Pesquisa – O “*Blog Auxílio à Pesquisa*” foi criado em maio de 2010, no começo, chamado de “*Blog Normalizando Documentos IPPUR*” foi o primeiro Blog da biblioteca. Inicialmente o *Blog* objetivava fornecer orientação à normalização de trabalhos acadêmicos para os alunos do Instituto de Pesquisa e Planejamento Urbano e Regional da Universidade Federal do Rio de Janeiro (IPPUR/UFRJ).

No início o *Blog* foi constituído de material proveniente das questões levantadas pelos alunos durante os momentos de orientação dos alunos basicamente um FAQ - Frequently Asked Questions. Os FAQs reúnem as respostas às perguntas mais comuns que os clientes fazem - e por isso foi construído no esquema de perguntas e respostas e com caráter estático, ou seja, não se permite comentários, para evitar possíveis ruídos na comunicação com os alunos. Deste modo, o *Blog* complementava as orientações de normalização dos trabalhos acadêmicos fornecidas pelo manual utilizado na época. Atualmente o *Blog* ampliou a orientação e oferece também informações gerais sobre o processo de pesquisa. Histórico de todas as visualizações de página totaliza atualmente 230.768 visualizações tanto no Brasil quanto no exterior, conforme ilustração abaixo:

Figura 1: Histórico de todas as visualizações do Blog Auxílio à Pesquisa



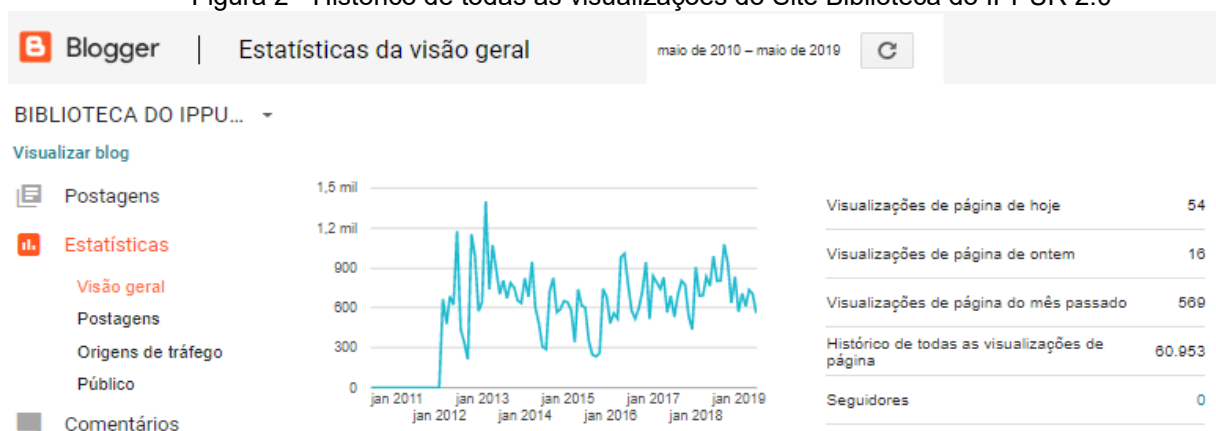
FONTE: Ferramenta de estatística do Blogger

Guia Prático para Normalização de Trabalhos Acadêmicos da Biblioteca do IPPUR/UFRJ – Sucessor do Manual de Elaboração de Trabalhos Acadêmicos do IPPUR, o *Guia Prático* foi elaborado pelos bibliotecários e lançado em novembro de 2011. Trata-se de uma obra totalmente ilustrada, estruturada de forma pedagógica e que condensa as normas da ABNT, além de ensinar, de uma maneira clara e direta, o processo de normalização dos trabalhos. Para manter o Guia Prático *Online* a biblioteca utiliza a ferramenta *Google Docs* que disponibiliza o arquivo digital em formato pdf do *Guia*. O objetivo é facilitar o download a obra. Desde 10 de março de 2019 (data da última atualização o *Guia* alcançou 3.366 visualizações de página.

Boletim de Novas Aquisições – Editado bimestralmente, o *Boletim de Novas Aquisições* arrola os itens que passaram a integrar o acervo da Biblioteca do IPPUR por meio de doações, permutas e compras. O *Boletim* traz a referência e o nº de chamada dos itens para facilitar a localização de cada publicação no seu acervo. Lançado em março de 2012, a sua versão digital da obra mantém a estrutura original da publicação em papel. O *Boletim* utiliza a ferramenta para desenvolver Blogs, Blogger. Desde sua criação o site teve 11.210 visualizações.

Site Biblioteca do IPPUR 2.0 – Lançado em março de 2012, o *Site Biblioteca do IPPUR 2.0* foi idealizado primeiramente com intuito de facilitar o atendimento ao usuário/cliente. Nossa preocupação inicial era criar uma marca, ou seja, uma identidade institucional na *Internet* (ciberespaço) para a biblioteca. Esse trabalho já está em andamento em órgãos públicos e privados preocupados em aumentar a visibilidade de suas instituições, mas pouco se ler sobre essa prática para bibliotecas que se fazem presente no mundo virtual. Assim, foi criado um ambiente *Web* que aglutinava todos os produtos desenvolvidos pela biblioteca num só lugar. Em um só endereço de *Web*. Isso porque acreditávamos que gerir um site próprio independente dos portais institucionais poderia valorizar as bibliotecas junto ao seu público. Desde sua criação, até maio de 2019, o site teve 60.953 visualizações.

Figura 2 - Histórico de todas as visualizações do Site Biblioteca do IPPUR 2.0



FONTE: Ferramenta de estatística do Blogger

Redes Sociais – Em bibliotecas, essas tecnologias de mídia social costumam ser utilizadas tanto para compartilhar notícias gerais sobre a biblioteca como para fornecer informações de algum assunto específico. Assim, a Biblioteca do IPPUR criou perfis nas redes sociais *Facebook* e *Twitter* com intuito de disponibilizar clipping diários com informes e notícias nas áreas: de cultura, biblioteconomia, Rio de Janeiro, planejamento urbano e regional, informações gerais sobre a UFRJ entre outros temas.

Criação, manutenção e monitoramento: conselhos úteis

A utilização das mídias sociais, no ambiente biblioteconômico universitário, pode contribuir tanto para a implantação (ou para o aperfeiçoamento) de produtos e/ou serviços oferecidos, como para facilitar a comunicação das bibliotecas com seus usuários. A manutenção da reputação da biblioteca nas redes sociais é de responsabilidade do administrador da conta. Manter a reputação significa, principalmente, controlar a informação postada. Desta forma:

- tenha consciência de que o usuário é o protagonista e que sua página existe para satisfazer primeiramente as necessidades dele, se alcançar outras pessoas será uma consequência do bom trabalho realizado;
- documente todas as atividades nas mídias sociais, para que a conta não pare de funcionar (inclusive as senhas de acesso);
- nomeie mais de um administrador;
- preocupe-se com a imparcialidade;
- poste sempre colocando títulos breves junto com textos e URL;
- adote um tom cordial quando responder às mensagens. Lembre-se de que toda mensagem deve ter direito de receber resposta.
- trate a página da biblioteca de forma diferente de um perfil pessoal, pois é um ato institucional;
- aprenda com outras instituições (siga-as e inspire-se em boas práticas);
- faça manutenção contínua da página. Poste de duas a quatro postagens por dia;
- mostre o que a sua biblioteca faz e também o que ela tem, ou seja, publique materiais relativos à: rotina da biblioteca, apresentação de itens do acervo, apresentação de produtos e serviços etc.;
- considere que existe muito material de boa qualidade produzido na Web. Avalie, poste, mas sempre identifique as fontes;
- divulgue o que é produzido em sua instituição (eventos, treinamentos, palestras, workshops, cartazes, folhetos, etc);
- monitore regularmente a página para saber se os objetivos estão sendo alcançados.

Uma boa forma de medir o desempenho na mídia social é monitoramento das atividades na página. Lembre-se de que quando você está medindo e avaliando o sucesso na mídia social é importante medir o que importa, ou seja, o que foi identificado inicialmente como propósito no processo de planejamento. Passado algum tempo de presença nas mídias, torna-se necessário as instituições compreender o desempenho das páginas na *Web* social. Comumente as mídias sociais disponibilizam para os administradores uma seção contendo o registro das atividades nas páginas sociais.

A avaliação do desempenho da utilização das mídias sociais da biblioteca consiste basicamente em examinar algumas métricas definidas como chave. Nesta seção, propomos um conjunto de métricas simples que a biblioteca acadêmica pode usar para medir o impacto de suas atividades. Os dados colhidos pela interface podem permitir que às BUs transformem essa ferramenta em um canal de marketing estratégico. Para isso, o profissional deve ser capaz de identificar os métodos de medição dos resultados desejados de uma mídia social. Para efeito de estudo do monitoramento recomendamos uma atenção especial para indicadores que consideramos relevantes nas mídias sociais *Facebook* e *Twitter* para revelar o desempenho do conteúdo de uma página. Esses indicadores estão presentes na interface de estatísticas das principais mídias sociais:

FACEBOOK

- INDICADOR 1:** Fãs alcançados
- INDICADOR 2:** Alcance orgânico
- INDICADOR 3:** Envolvimento
- INDICADOR 4:** Envolvimento dos visitantes
- INDICADOR 5:** Taxa de cliques

TWITTER

- INDICADOR 1:** Atividade do *Tweet*
- INDICADOR 2:** Contagem de seguidores
- INDICADOR 3:** Interesses dos seguidores
- INDICADOR 4:** Localização dos seguidores
- INDICADOR 5:** Sexo dos seguidores

Considerações finais e perspectivas futuras

Na literatura profissional encontram-se alguns estudos com os mais variados enfoques: padrões de comportamentos, opiniões dos bibliotecários relativas ao tema, além de análises sobre o desenvolvimento de uma mídia social específica. A motivação deste estudo que se justifica em primeiro lugar, por permitir um compartilhamento de experiência e também por contribuir para um melhor entendimento do uso dessas ferramentas.

Acreditamos que estudar as mídias sociais traz benefícios para os esforços de marketing de bibliotecas acadêmicas. Benefícios estes, óbvios para as organizações comerciais e ainda nebulosos para as bibliotecas. O relacionamento dos indivíduos com as mídias sociais é um assunto relativamente recente e rico, mas que ainda carece de muitos estudos. Prever o futuro da mídia social é uma tarefa muito difícil. O futuro das bibliotecas acadêmicas, neste universo contínuo e onipresente, está aberto ao debate.

Nessa perspectiva, ganha destaque a necessidade de bibliotecas acadêmicas serem mais estratégicas, para divulgar seus serviços e/ou produtos. E a mídia social oferece uma oportunidade para esta iniciativa. Concluindo podemos dizer que as ferramentas 2.0 se revelaram um ótimo instrumento para a biblioteca, uma excelente vitrine para a exposição de seus produtos e serviços. Vale ressaltar, também, que as ações desenvolvidas:

- melhoram a comunicação internamente e externamente, ou seja, o compartilhamento do conhecimento dentro da organização de forma mais ampla (universidade, instituto, centro, etc.), e também no local de trabalho (a própria biblioteca);
- facilitaram a criação de serviços personalizados, pois nos dias atuais as BUs não têm fronteiras e costumam ser avaliadas em função dos serviços que prestam e não

pela dimensão de suas coleções, ou seja, agora a biblioteca é avaliada pelo que faz e não pelo que ela tem;
➤ além de atrair novos usuários.

Referências

ANJOS, Cláudia Regina dos. **Mídias sociais nas bibliotecas da UFRJ**: adoção e monitoramento. 2016. 162f. Dissertação (Mestrado em Biblioteconomia) - Programa de Pós-graduação em Biblioteconomia, Centro de Ciências Humanas e Sociais, Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2016. Orientador: Moreno Albuquerque de Barros.

ARAÚJO, Felipe. **Leis de Ranganathan**. Disponível em:
<<http://www.infoescola.com/curiosidades/leis-de-ranganathan/>>. Acesso em: 10 out. 2012. INFOESCOLA: navegando e aprendendo.

BARROS, Moreno Albuquerque de. **Emergência e dinâmica informacional na Blogosfera**. Dissertação – Programa de Pós-graduação em Ciência da Informação. Mestrado em Ciência da Informação, Convênio Universidade Federal Fluminense - Instituto de Arte e Comunicação Social, Instituto Brasileiro de Informação em Ciência e Tecnologia. Niterói, Rio de Janeiro. 2009.

BEHR, Ariel; MORO, Eliane L. S.; ESTABEL, Lizandra B. Uma proposta de atendimento às necessidades de informação dos usuários da biblioteca escolar por meio do benchmarking e do sensemaking. **Informação & Informação**, Londrina, v. 15, n. 1, p. 37 - 54, jul./jun. 2010.

BEHR, Ariel et al. Especialização em bibliotecas escolares e acessibilidade: discutindo a gestão da biblioteca na modalidade EAD. **Informação & Informação**, Londrina, v. 16, n.1, p. 102-123, jan./jun. 2011. Disponível em:
<<http://www.uel.br/revistas/uel/index.php/informacao/article/view/7831/8481>>. Acesso em: 26 set. 2011.

CALIL JUNIOR, Alberto; CORRÊA, Elisa Cristina Delfini; SPUDEIT, Daniela. O uso das mídias sociais nas bibliotecas brasileiras: análise dos trabalhos apresentados no SNBU e CBBB. In: Congresso Brasileiro de Biblioteconomia, Documentação e Ciência da Informação-FEBAB. 2013 **Anais**. p. 5044-5059. Disponível em: <<http://portal.febab.org.br/anais/article/view/1634/1635>>. Acesso em: set. 2015.

CUNHA, Murilo Bastos da. A biblioteca universitária na encruzilhada. **DataGramZero**: Revista de Ciência da Informação, Rio de Janeiro, v. 11, n.6, dez. 2010. Disponível em: <http://www.datagramazero.org.br/dez10/Art_07.htm>. Acesso em: 6 fev. 2015.

FIDELIZAÇÃO de clientes: 5 dicas essenciais para encantar o comprador manter o cliente pode ser mais importante do que conquistar um novo. 26 jun. 2015. **Revista**

Pequenas Empresas & Grandes Negócios. Disponível em: <http://revistapegn.globo.com/Dia-a-dia/noticia/2015/03/fidelizacao-de-clientes-5-dicas-essenciais-para-encantar-o-comprador.html>. Acesso em: 15 jul. 2016.

FRIZON, Jaqueline. **Geração C:** os futuros profissionais: como as empresas estão se adequando para receberem a turma que não vive sem as mídias sociais. 2015. Disponível em: < <http://carreiraenegocios.uol.com.br/gestao-motivacao/41/artigo255115-2.asp>>. Acesso em: 3 fev. 2015.

JESUS, Deise Lourenço de; CUNHA, Murilo Bastos da. Produtos e serviços da *Web 2.0* no setor de referência das bibliotecas. **Perspectivas em Ciência da Informação**, v. 17, n. 1, p. 110-133, jan./mar. 2012. Disponível em: <http://www.scielo.br/scielo.php?pid=S1413-99362012000100007&script=sci_arttext>. Acesso em: 30 jan. 2015.

KEATING, Roly. Rise of smart phones and social media makes people crave real books. **The Telegraph**, 25 June 2015. Disponível em: <<http://www.telegraph.co.uk/news/uknews/11343797/Rise-of-smart-phones-and-social-media-makes-people-crave-real-books-British-Library-report-finds.html>>. Acesso em: 23 jun. 2015. Entrevista concedida a Hannah Furness.

KOTLER, Philip. **Administração de marketing**. 10. ed. São Paulo: Prentice Hall, 2000.

NASCIMENTO, Cecília Maria Pereira. **A biblioteca universitária hoje:** gerencia compartilhada. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 2000. Disponível em: <<http://snbu.bvs.br/snbu2000/docs/pt/doc/t026.doc>>. Acesso em: 03 jul. 2005.

A PESQUISA atualiza, inova e qualifica as atividades acadêmicas, diz reitor da UFRJ. In: ENCONTRO INTERNACIONAL DE REITORES UNIVERSIA, 3. Discurso de Carlos Levi, Reitor da UFRJ. Disponível em: < <http://noticias.universia.com.br/destaque/noticia/2014/07/28/1101344/a-pesquisa-atualiza-inova-qualifica-atividades-academicas-diz-reitor-ufrrj.html>>. Acesso em: set 2014.

PRADO, Heloísa de Almeida. **Organização e administração de bibliotecas**. 2. ed. São Paulo: T. A. Queiroz, 1992.

QUESADA, Beatriz. Desafios para tornar as bibliotecas públicas mais atraentes: além de problemas estruturais, imagem negativa do ambiente afasta o público. **Revista Educação**. Extras. Edição 230. Disponível em: < <http://revistaeducacao.com.br/textos/230/desafios-para-tornar-as-bibliotecas-publicas-mais-atraentes-371778-1.asp>>. Acesso em: 15 jul. 2016.



VIANA, Michelângelo Mazzardo Marques. **A informação e a Biblioteca Universitária**. 2013. Disponível em: < <http://pt.slideshare.net/miquemv/ss-a-informao-e-a-biblioteca-universitria> >. Acesso em: 24 jan. 2015.

Acesso aberto, dados abertos e ciência aberta no Brasil: revisão sistemática de literatura a partir do Catálogo de Teses e Dissertações da CAPES

Débora Sarmento Carvalho¹, Marta Macedo Kerr Pinheiro²

¹ Mestranda no Programa de Pós-Graduação em Sistemas de Informação e Gestão do Conhecimento na FUMEC (MG). Email: Diadorim.lua@gmail.com. ORCID: 0000-0001-5172-3864

² Professora doutora do Programa de Pós-Graduação em Sistemas de Informação e Gestão do Conhecimento na FUMEC (MG). Email: martapinheiro@ufmg.br. ORCID: 0000-0001-5592-3396

Resumo

A discussão sobre o acesso aberto foi ampliada pelo debate em torno da disponibilização dos dados abertos de pesquisa, levando a repensar a própria prática científica como proposto pela Ciência Aberta. Essas novas práticas se alinham à estrutura de rede, mais dinâmicas, horizontais e colaborativas, em que se compartilha tanto dados quanto ferramentas. O presente trabalho propôs o mapeamento e sistematização da literatura especializada sobre esses temas, utilizando o Catálogo de Teses e Dissertações da Capes. Foram recuperados 121 trabalhos que abordam o acesso, dados e ciência aberta sob os mais diversos aspectos – direito autoral, implantação de repositórios, usabilidade, estudo de usuários, políticas de informação, formato de dados e transparência, dentre outros. Os repositórios institucionais são o principal foco dos trabalhos disponíveis na Plataforma (25), seguidos das ferramentas para publicação/consumo de dados abertos (13). Entretanto, as discussões que relacionam os três conceitos mostram-se pontuais e reforçam a necessidade de aprofundamento nos estudos e, sobretudo, nas práticas defendidas por esses movimentos.

Palavras-chave: Acesso aberto. Dados Abertos. Ciência Aberta. Revisão Sistemática de Literatura.

Abstract

The discussion about open access was expanded by the debate about the availability of open research data, leading to rethinking the scientific practice itself as proposed by Open Science. These new practices are aligned to the network structure, more dynamic, horizontal and collaborative, in which both data and tools are shared. The present work proposed the mapping and systematization of the specialized literature on these subjects, using the Catalog of Thesis and

Dissertations of Capes. A total of 121 papers covering access, data and open science under different aspects - copyright, repository deployment, usability, user study, information policies, data format and transparency, among others, were retrieved. Institutional repositories are the main focus of the works available in the Platform (25), followed by tools for publication / consumption of open data (13). However, the discussions that relate the three concepts are punctual and reinforce the need to deepen the studies and, above all, the practices defended by these movements.

Keywords: Open access. Open Data. Open Science. Systematic Review of Literature.

Introdução

Desde 2002, com a Declaração de Budapeste, marco do Movimento *Open Access*, há uma crescente discussão sobre a comunicação científica e à própria prática da ciência. Correlato ao *Open Access*, o movimento Ciência Aberta defende a livre circulação de bens culturais, sendo marcado pelo desenvolvimento de formas de apropriação da informação baseadas na liberdade de (re)produção, acesso e uso de conteúdos intelectuais. Essas novas práticas se alinham à estrutura de rede, mais dinâmicas, horizontais e colaborativas, em que se compartilha tanto dados quanto ferramentas. (Cardoso&Jacoberty, s.d)

Muitos estudiosos têm-se debruçado sobre as questões relativas à essas novas propostas de produção e consumo de saberes. Com o intuito de mapear e sistematizar o conhecimento disponível na literatura especializada sobre dados abertos, acesso aberto e ciência aberta de forma a evidenciar os esforços da comunidade acadêmica nacional, especialmente no que diz respeito ao volume de artigos e conceitos utilizados, foi realizada uma revisão sistemática utilizando o Catálogo de Teses e Dissertações da Capes. Foram recuperados 121 trabalhos que abordam as questões listadas sob os mais diversos aspectos – direito autoral, implantação de repositórios, usabilidade, políticas de informação, formato de dados e transparência, dentre outros.

Acesso aberto

Com o advento da Internet, o compartilhamento de informações entre os cientistas adaptou-se ao ambiente digital e a estrutura reticular do novo ambiente potencializou a prática das trocas espontâneas entre os cientistas. (Swan&Brown, 2005)

Projetos e iniciativas de autoarquivamento e de conteúdo de acesso aberto se multiplicaram e, em 1999, responsáveis por repositórios de versões digitais de publicações científicas se reuniram na Convenção de Santa Fé, Novo México. Ainda que a tônica do evento tenha sido os aspectos técnicos, ele influenciaria o surgimento de ações politizadas acerca da divulgação científica, que foram sistematizadas no Movimento Open Access (Santos, 2014).

As iniciativas de Budapeste e Bethesda, em 2002, formataram o conceito de Acesso Aberto tal como se reconhece hoje. Os documentos recomendam a disponibilização online e sem limitações dos resultados dos trabalhos científicos, independentemente de haver ou não revisão por pares. Cabe observar que esses documentos tinham como pauta central a forma de compartilhamento do conhecimento científico, ou seja, não tratam da produção do saber (Cardoso, 2009). Com a Declaração de Berlim (2003), há avanços que vão iniciar a discussão sobre uma nova forma de fazer científico. Nela, são listados os tipos de documentos que devem ser de acesso aberto: além dos resultados, devem estar disponíveis dados não processados, metadados, fontes originais, dentre outros, garantindo acesso às fontes primárias e secundárias das pesquisas.

Em 2007, os Princípios e Diretrizes para Acesso aos Dados de Pesquisa de Financiamento Público, da Organização para a Cooperação e Desenvolvimento (OCDE), reforça a ampliação da discussão do acesso aberto para dados de pesquisa abertos (Chignard, 2013). O documento defende que o acesso aos dados aumenta o retorno dos investimentos públicos em ciência e tecnologia, incentiva a diversidade de estudos permitindo, inclusive, a exploração de tópicos não imaginados pelos investigadores iniciais.

Esse contexto levou a um novo patamar e o debate se expande do compartilhamento de resultados para cada uma das fases do ciclo de vida da pesquisa (Caruso, 2015). A ciência aberta defende uma metodologia de construção coletiva do conhecimento, democratizando o acesso e contribuindo para o desenvolvimento, inclusão social e formação cidadã. O acesso à informação científica pode contribuir para uma participação mais qualificada e efetiva dos cidadãos, especialmente no contexto em que se considera a necessidade de estabelecer políticas públicas baseadas em evidências (Willinsky, 2006; Santos, 2017).

A ciência aberta argumenta que os mecanismos de restrições atrasam o avanço do conhecimento, impactando o desenvolvimento econômico e social. A liberação de uso e circulação, por sua vez, poderia incentivar a exploração e a reutilização por outros pesquisadores, assim como incentivar o trabalho cooperativo na resolução de determinados problemas científicos (Santos, 2014; Costa, 2006) de forma mais global, mais rápida e efetiva, e ainda com menores custos (Andrade, 2014).

No aspecto econômico, a Ciência Aberta advoga a necessidade de acesso ao conteúdo resultante de pesquisa financiada com verba pública a partir do entendimento de que a sociedade já pagou por ele através do recolhimento dos impostos. Nesse sentido, mais do que o compartilhamento motivado pela noção de que o conhecimento é um bem público, o compartilhamento e a colaboração ganham um aspecto mais político no sentido de uma transparência que, entre outros, beneficia a sociedade (Houcarde, 2015; Caruso, 2015).

Metodologia

Com o intuito de evidenciar a inserção da comunidade acadêmica no debate da ciência aberta, foi realizado um mapeamento e sistematização do conhecimento disponível na literatura especializada sobre dados abertos, acesso aberto e ciência aberta. Optou-se por estudar os recursos disponíveis no Catálogo de Teses e Dissertações da Capes (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - <http://catalogodeteses.capes.gov.br>), que reúne registros bibliográficos desde 1987. Tanto os metadados quanto os arquivos completos dos trabalhos de mestrado e doutorado de programas reconhecidos são disponibilizados na Plataforma Sucupira e sincronizados periodicamente com o Catálogo.

A Plataforma Sucupira é utilizada para coletar informações, tendo como objetivos principais maior transparência dos dados para toda a comunidade acadêmica, redução de tempo, esforços e imprecisões na execução de avaliação do Sistema Nacional de Pós-graduação, maior confiabilidade, precisão e segurança das informações.

A disponibilização de teses e dissertações produzidas pelos programas de pós-graduação stricto sensu tem como referência a Portaria 13/2006. Ela instituiu a regra a partir das manifestações do Conselho Técnico-Científico, verificadas em 2005, de que a produção discente é um indicador da qualidade dos programas e que não poderia ser aferida somente através da publicação seletiva nos periódicos especializados.

Em decorrência da obrigatoriedade estabelecida pela Portaria, os programas deveriam manter repositórios próprios ou utilizar um sistema recomendado pela Capes, que, por sua vez, deveria divulgar a lista dos artigos em um ambiente digital online adequado.

Por essas características, o Catálogo de Teses e Dissertações da Capes foi escolhido para verificar o estado da arte das discussões sobre dados abertos e acesso aberto na produção científica brasileira.

Para a realização da revisão sistemática da literatura foi adotado o processo proposto por Kitchenham (2004) que se divide em três etapas distintas: Planejamento da Revisão, Condução da Revisão e Análise dos Resultados.

Planejamento da revisão

Na fase de planejamento foi definido um protocolo para a execução da Revisão Sistemática da Literatura que seguiu as etapas descritas na sequência:

- a) Descrição dos objetivos: identificar artigos que abordam conceitos relativos a acesso aberto, dados abertos e ciência aberta
- b) Elaboração das questões de pesquisa:
 - quantos trabalhos são relacionados aos conceitos de acesso aberto, dados abertos e ciência aberta?
 - quais as relações estabelecidas entre os temas?

- quais são os pesquisadores e centros de pesquisa em se tratando dessa temática?
- quais as áreas de conhecimento e áreas de avaliação dos estudos?
- quantos e quais os trabalhos monográficos foram divulgados no formato de artigo em periódico revisado por pares?

c) A estratégia de busca utilizada neste trabalho baseia-se em dois critérios:

- seleção de fonte de pesquisa: utilizou-se a base de dados do Catálogo de Teses e Dissertações da Capes para obtenção dos estudos.
- elaboração da *string* de busca: foi elaborada uma *string* de busca que levou em consideração os termos “acesso aberto”, “open access”, “dados abertos”, “open data”, “ciência aberta” e “open Science”.

d) Adoção de critérios para a inclusão de exclusão dos trabalhos selecionados:

- para a inclusão dos estudos: as publicações devem ter os textos completos disponíveis na base citada, em português, e devem responder a qualquer uma das questões de pesquisa.
- critérios para a exclusão dos estudos: foram excluídas as teses e dissertações que não abordavam os conceitos de acesso, dados e ciência aberta no contexto estudado, qual seja, da cultura aberta.

Condução da revisão sistemática

A condução da revisão sistemática observou os passos: as *strings* de busca foram executadas nas fontes selecionadas; os estudos primários foram identificados; os documentos retornados foram selecionados de acordo com os critérios de inclusão/exclusão e avaliados segundos os critérios de qualidade estabelecidos durante o planejamento.

Processo para recuperação de pré-seleção dos estudos primários

Nesta etapa foram recuperados os trabalhos preliminares e realizado um processo de filtro nos estudos. Foram encontrados 283 estudos, a partir das *strings* definidas na etapa de planejamento. Elas foram, posteriormente, agrupadas para análise, de forma que os termos “dado aberto”, “dados abertos” e “open data” são referidos apenas como “dados abertos”; as *strings* “acesso aberto” e “open access” são tratadas como “acesso aberto” e, por sua vez, “ciência aberta” e a sua variação em inglês foram agrupadas em torno do primeiro termo. Assim, ao serem consolidados, os números de trabalho referentes a cada termo são diferentes do somatório das *strings* individuais, já que um mesmo texto foi recuperado por mais de uma *string* (quadro 1).

Além disso, três trabalhos relacionam dados abertos e acesso aberto e dois o fazem em relação aos conceitos acesso aberto e ciência aberta. Assim, dos 288 trabalhos resultantes da soma dos resultados acima, reduziu-se 5 trabalhos que figuravam em mais de um termo, chegando aos 283 estudos. Após verificar a disponibilidade, chegou-se a 150 trabalhos que tinham os textos completos disponíveis no Catálogo de Teses e Dissertações. Adotando o critério idiomático, reduziu-se outros oito estudos, chegando ao volume inicial de 142 teses e dissertações. Outros 56 trabalhos foram incluídos na amostra por serem anteriores à Plataforma Sucupira, sendo um deles em inglês, restando 55 trabalhos. Um outro trabalho foi incluído porque havia um documento disponível na plataforma indicando o caminho para outro repositório em que o trabalho podia ser acessado. Desta forma, encerrou-se a etapa de recuperação e seleção de estudos com 198 textos.

Quadro 1 – organização dos estudos por temas e agrupamentos

Dados abertos	<p>dados aberto (03)</p> <p>dados abertos (79)</p> <p>open data (92)</p>	113
Acesso aberto	<p>acesso aberto (94)</p> <p>open access (116)</p>	156
Ciência aberta	<p>ciência aberta (17)</p> <p>open Science (10)</p>	19

Fonte: elaboração própria

Processo de seleção dos estudos

Após a recuperação e pré-seleção, os artigos recuperados foram analisados pela leitura do seu conteúdo. Dessa forma, os 198 documentos recuperados foram analisados e, destes, foram selecionados 121 artigos, considerando apenas aqueles que tratavam dos conceitos abordados no presente trabalho.

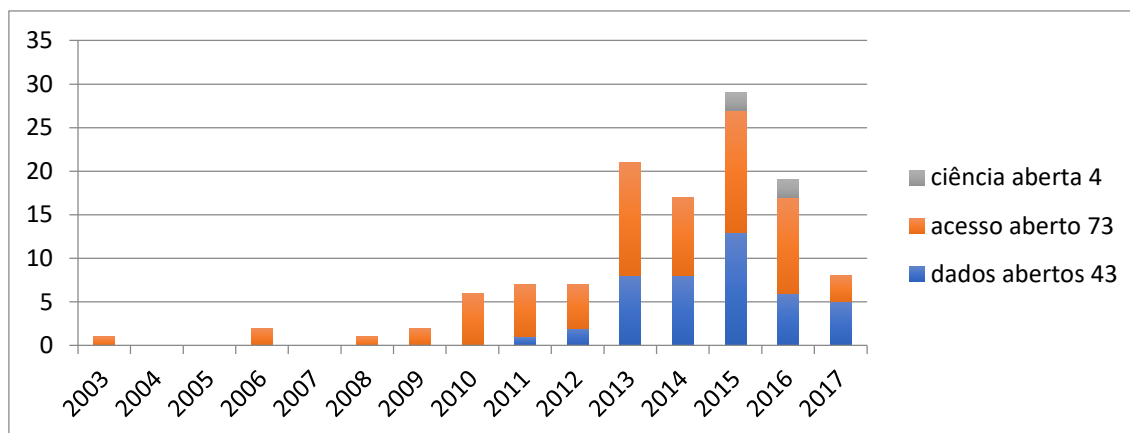
Resultados

Os estudos de dados abertos e acesso aberto se concentram nos mestrados acadêmicos, variando entre 60% e 70% dos estudos produzidos. A ciência aberta, por sua vez, tem uma divisão mais igualitária, sendo 53% dos estudos inseridos no mestrado acadêmico e 47% no doutorado. Essa distribuição é mantida ao se analisar a disponibilidade na plataforma, o idioma escolhido e os conceitos relativos à acesso, dados e ciência aberta.

Pode-se observar que os estudos acompanham a tendência de discussão mundial, iniciando em 2003 (uma trabalho disponível), incluindo as discussões sobre dados abertos a partir de 2011, coincidindo com os debates em da Lei 12.527/2011, conhecida como Lei de Acesso à Informação (um estudo disponível)

e, mais recentemente, em 2015, podem ser observados dois trabalhos recuperados à partir da *string* ciência aberta (gráfico 1)¹:

Gráfico 1: Evolução dos trabalhos sobre acesso aberto, dados abertos e ciência aberta

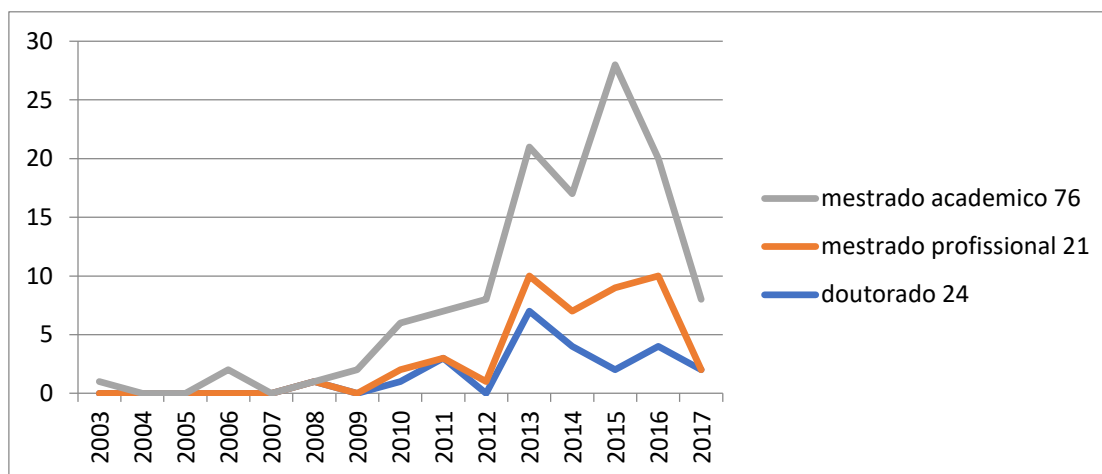


Fonte: elaboração própria

De forma similar, observa-se que a distribuição dos estudos sobre os três temas, quando divididos por níveis acadêmicos (mestrado acadêmico, mestrado profissional e doutorado), têm um desenho semelhante, com aumento notável no período de 2013 a 2015 e, a partir de então, uma tendência de queda, conforme gráfico 2.

¹ Em relação à ciência aberta, cabe observar que outros trabalhos, ainda no final da década de 1990, mas incluindo outros da década de 2010, relacionam o termo à característica mutável da ciência, referindo-se à possibilidade de outras discussões e entendimentos. Em relação ao acesso aberto, sem corte temporal específico, observa-se o termo sendo utilizado para se referir ao acesso físico a locais ou, na área da saúde, acesso face-a-face. Esses trabalhos foram descartados no processo de seleção dos estudos, mas são interessantes no sentido de se compreender que os conceitos de acesso aberto e ciência aberta ainda estão em disputa no território acadêmico.

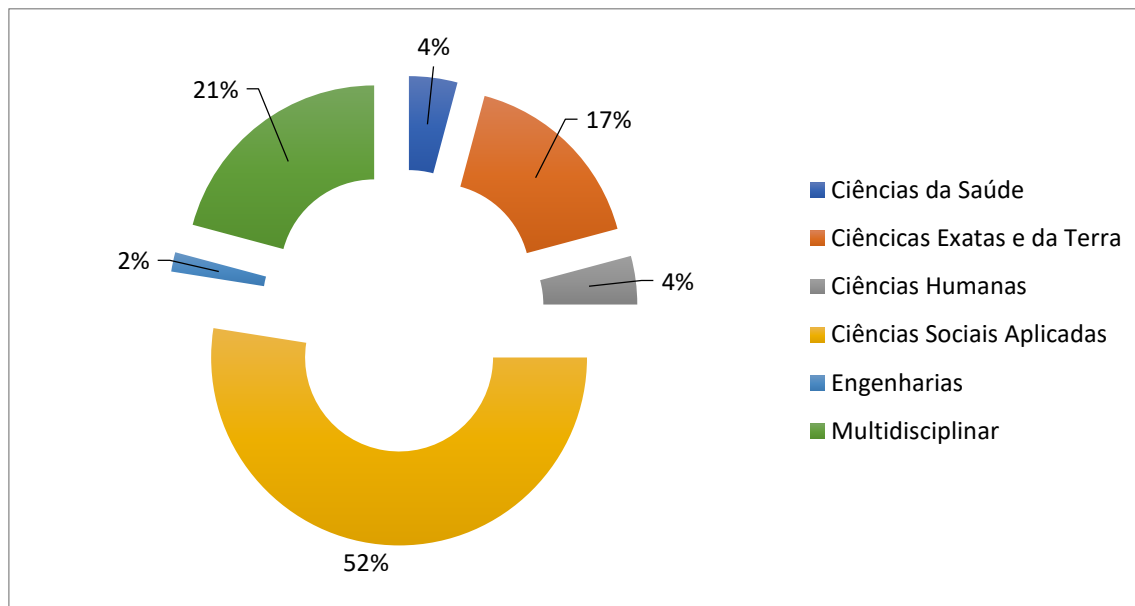
Gráfico 2: Teses e Dissertações sobre dados abertos, acesso aberto e ciência aberta (2003-2017)



Fonte: elaboração própria

A área de Ciências Sociais Aplicadas é a que concentra o maior número de trabalhos, o que pode ser explicado pelos debates acerca da disponibilização dos documentos (repositórios e periódicos), comunicação da ciência e transparência pública. A área multidisciplinar, no que tange à articulação entre o debate teórico e a prática da cultura aberta, está em segundo lugar, seguida das Ciências Exatas e da Terra, que enfatizam o desenvolvimento de ferramentas para garantir os dados e o acesso aberto (gráfico 3).

Gráfico 3: Estudos divididos por Área de Conhecimento



Fonte: elaboração própria

Os repositórios institucionais são o principal foco dos trabalhos disponíveis na Plataforma (25), seguidos das ferramentas para publicação/consumo de dados abertos (13) (gráfico 4). A preocupação em constituir, povoar e manter os repositórios pode ser entendida no sentido de se dar publicidade à produção de cada uma das instituições, garantindo uma boa reputação e maior reconhecimento entre os pares. Já em relação às ferramentas de publicação/consumo dos dados abertos, há uma preocupação em criar possibilidades para o uso dos dados pelos cidadãos, já que, em muitos trabalhos se argumenta que a disponibilização dos dados não é garantia de transparência ou que eles vão ser usados pelos cidadãos, já que, no formato aberto, requerem alguma competência técnica para que sejam manipulados.

Gráfico 4: distribuição temática dos trabalhos



Fonte: elaboração própria

Ao todo, 36 instituições de ensino disponibilizaram teses e dissertações pelo Catálogo da Capes. Dessas, apenas 10 tiveram pelo menos cinco trabalhos listados. Apesar desses números, não é possível afirmar que não existam outros polos de pesquisa sobre as temáticas abordadas na presente revisão de literatura, uma vez que, apesar de se propor a unificar a produção científica nacional, listando pelo menos os resumos do material disponível, o Catálogo da Capes apresenta descompasso em relação a outros repositórios, notadamente os das instituições listadas na tabela 1.

Tabela 1: Lista de Instituições de Ensino com trabalhos disponíveis no Catálogo

UNIVERSIDADE FEDERAL DE SANTA CATARINA	14
UNIVERSIDADE DE BRASÍLIA	13
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO	11
UNIVERSIDADE FEDERAL DE SÃO CARLOS	8
UNIVERSIDADE EST. PAULISTA JÚLIO DE MESQUITA FILHO/MARILIA	6
UNIVERSIDADE FEDERAL DO PARANÁ	6
UNIVERSIDADE DE SÃO PAULO	6
UNIVERSIDADE FEDERAL FLUMINENSE	6
PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO	5
UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO	5
UNIVERSIDADE FEDERAL DA BAHIA	4
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL	3
FUNDAÇÃO OSWALDO CRUZ	3
UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ	2
UNIVERSIDADE FEDERAL DE CAMPINA GRANDE	2
UNIVERSIDADE DO VALE DO RIO DOS SINOS	2
UNIVERSIDADE FEDERAL DE SÃO PAULO	1
FUNDAÇÃO UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL	1
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE	1
FUNDAÇÃO UNIVERSIDADE DE PASSO FUNDO	1
UNIVERSIDADE ESTADUAL DE LONDRINA	1
UNIVERSIDADE CATÓLICA DE BRASÍLIA	1
FACULDADE DE TECNOLOGIA SENAI CIMATEC	1
CENTRO DE FORMAÇÃO, TREINAMENTO E APERFEIÇOAMENTO	1
UNIVERSIDADE DE FORTALEZA	1
FACULDADE DE MEDICINA DO ABC	1
UNIVERSIDADE DO ESTADO DE SANTA CATARINA	1
UNIVERSIDADE FEDERAL DE SANTA MARIA	1
UNIVERSIDADE FEDERAL DO CEARÁ	1
UNIVERSIDADE ESTADUAL DO CEARÁ	1
UNIVERSIDADE FEDERAL DE MINAS GERAIS	1
UNIVERSIDADE FEDERAL DO AMAZONAS	1
UNIVERSIDADE FEDERAL DE PERNAMBUCO	1
UNIVERSIDADE FEDERAL DA PARAÍBA/JOÃO PESSOA	1
UNIVERSIDADE ESTADUAL DE CAMPINAS	1
UNIVERSIDADE FUMEC	1

Fonte: elaboração própria

Considerações

O presente trabalho propôs o mapeamento e sistematização da literatura especializada sobre esses temas, utilizando o Catálogo de Teses e Dissertações da Capes. Foram recuperados 283 trabalhos que abordam as questões listadas sob os mais diversos aspectos – direito autoral, implantação de repositórios, usabilidade, estudo de usuários, políticas de informação, formato de dados e transparência, dentre outros. Pode-se observar que os estudos acompanham a tendência de discussão mundial, iniciando em 2003, incluindo os dados abertos a partir de 2011, coincidindo com os debates em da Lei 12.527/2011, conhecida como Lei de Acesso à Informação e, mais recentemente, em 2015, podem ser verificados trabalhos que abordam a Ciência Aberta.

Inicialmente, a proposta desta revisão sistemática de literatura pretendia listar artigos derivados dos trabalhos apresentados. Entretanto, embora seja possível incluí-los entre os dados de cada trabalho através da Plataforma Sucupira, de forma que fossem recuperados pelo Catálogo de Teses e Dissertações da Capes, não se verificou sua utilização por parte dos pesquisadores, sendo descartada. Entretanto, considera-se que a limitação deste estudo, especialmente o tange ao volume de artigos disponíveis no Catálogo, assim como o que diz respeito à não utilização de recursos tais como indicação de fontes de financiamento, grupos de pesquisa e trabalhos derivados, originalmente previstos no sistema da Capes, podem motivar trabalhos futuros para entender a motivação ou não do preenchimento de tais campos.

A maioria dos trabalhos sobre acesso aberto focam os sistemas de informação para acesso aos resultados das pesquisas, sejam eles em artigos de periódicos científicos, sejam em repositórios institucionais. As questões técnicas (padrões e formatos, arquitetura de informação e metadados), questões gerenciais (política de informação, gestão do conhecimento) e jurídicas (direito autoral) também são temas de trabalhos, embora em volume menor que os dois primeiros.

É importante notar que, especificamente sobre a disponibilização dos dados abertos, há um debate sobre uma possível assimetria de informação gerada ou aprofundada por essa prática, uma vez que requer do usuário certo nível de conhecimento técnico para que faça uso do recurso. Nesses trabalhos, houve uma preocupação em desenvolver ferramentas que facilitassem o uso e reuso de dados de forma não-discriminatória, conforme preconizam as diretrizes de dados abertos.

Na análise geral das teses e dissertações, pode-se observar que, embora haja uma discussão sobre a comunicação científica e outra sobre dados abertos, em momentos muito pontuais elas convergem para o debate acerca do uso e reuso dos dados ou dos achados científicos comunicados em acesso aberto. Embora se discuta o acesso à produção científica, em poucos momentos há uma discussão sobre o próprio fazer da ciência e de que forma a adoção do acesso aberto e dos dados abertos contribuiriam para um novo fazer científico. Esses debates ficaram restritos aos trabalhos sobre ciência aberta.

Referências

- Andrade, V. T. A. A. (2014). Comunicação científica na sociedade em rede: uma plataforma de ciência aberta para o Brasil. 227 fl. (Tese de Doutorado em Comunicação). Centro de Artes e Comunicação. Universidade Federal de Pernambuco, Recife, PE.
- Cardoso, G. & Jacobetty, P. (sem data) O que significa Open Science? Lisbon Internet and Networks Institute. Recuperado a partir de: http://www.lini-research.org/np4/?newsId=12&fileName=open_science.pdf
- Cardoso, G., Caraça, J., Espanha, R., Triães, J., & Mendonça, S. (2009). As políticas de Open Access: Res publica científica ou autogestão?. *Sociologia, problemas e práticas*, (60), 53-67.
- Caruso, F. D. S. (2015). Ciência Aberta: Ações de Pesquisadores Acadêmicos na Web Aberta. (Dissertação de Mestrado em Ciência da Informação). Universidade Federal Fluminense, Niteroi, RJ.
- Chignard, S. (2013). *A brief history of open data*. *Paris Tech Review*, 29.
- Costa, S. M. D. S. (2006). Filosofia aberta, modelos de negócios e agências de fomento: elementos essenciais a uma discussão sobre o acesso aberto à informação científica. *Ciência da Informação*, Brasília, DF, v. 35, n. 2, p.39-50.
- Houcarde, V. (2015). O movimento Ciência Aberta no Brasil. 152 f. (Dissertação de Mestrado Divulgação Científica e Cultural). Universidade Estadual de Campinas, Campinas, SP.
- Kitchenham, B. (2004) "Procedures for Performing Systematic Reviews", Joint Technical Report Software Engineering Group, Keele University, United Kingdom and Empirical Software Engineering, National ICT Australia Ltd, Australia.
- Santos, J. C. F. D. (2014). Estudo sobre o movimento Open Access e de suas implicações para a comunicação na ciência. 144 fl. Dissertação (Mestrado em Política Científica e Tecnológica). Instituto de Geociências da Universidade Estadual de Campinas, SP.
- Santos, P. X. D., Almeida, B. D. A., Elias, F., Motta, M. L. D., Guanaes, P., Jorge, V. D. A., ... & Oliveira, G. (2017). *Livro Verde-Ciência aberta e dados abertos: mapeamento e análise de políticas, infraestruturas e estratégias em perspectiva nacional e internacional*.
- Swan, A., & Brown, S. (2005). *Open Access self-archiving: An author study*. 104p. UK FE and HE funding councils.
- Willinsky, J. (2006). *The access principle: The case for Open Access to research and scholarship*. Cambridge, Mass.: MIT Press.

Visibilidade da ciência latino-americana: contribuições da rede La Referencia para a ampliação do acesso aberto

Izabel Antonina de Araujo¹, Camila Mariana Aparecida da Silva², Luciana Gonçalves Silva Souza³

¹ Bibliotecária Universidade Federal de Minas Gerais e doutoranda pelo Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Minas Gerais. Email: antoninaizabel@gmail.com

² Bibliotecária Universidade Federal de Minas Gerais e doutoranda pelo Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Minas Gerais. Email: camilamas.silva@gmail.com

³ Bibliotecária Universidade Federal de Minas Gerais e doutoranda pelo Programa de Pós-Graduação em Gestão e Organização do Conhecimento da Universidade Federal de Minas Gerais. Email: lucianags9@gmail.com

Resumo

Nas últimas décadas o movimento pela democratização do conhecimento por meio do acesso aberto e da ciência aberta passou a ganhar maior visibilidade com a adoção de novas tecnologias e a formação de redes colaborativas. Neste sentido, a formação da Rede Federada de Repositórios Institucionais de Publicações Científicas (LA Referencia) constitui marco para a divulgação da produção científica da América Latina. O presente trabalho tem como objetivo analisar as ações desenvolvidas pela Rede LA Referência relacionadas à ampliação do acesso aberto nesta região a fim de conhecer e explicitar os elementos que subsidiam a formação de redes de repositórios. Para tanto, realiza pesquisa qualitativa de caráter descritivo que utiliza como técnicas de investigação a revisão de literatura e a análise documental de acordos políticos e técnicos, além de diretrizes estabelecidas pela Rede e deliberações do Conselho Diretivo. A proposta se justifica pela relevância de explorar iniciativas que, estando fora do eixo central da produção científica, possam se tornar referência na formação de novas redes de repositórios que almejem estabelecer políticas comuns e ampliar sua visibilidade em nível regional ou mundial. Como conclusão aponta que a iniciativa contribui para a economia de recursos, amplia a capacidade de atuação para divulgação da produção científica em âmbito internacional, além de aglutinar investimentos e força político-econômica em favor do acesso aberto.

Palavras-chave: acesso aberto; repositórios; LA Referencia; América Latina.

Introdução

Nos últimos anos, a criação de redes integradoras de repositórios tem ganhado notoriedade no cenário mundial. Estas redes, por meio de acordo de

cooperação, têm impulsionado a criação de um conjunto de políticas comuns que contribuem para a criação de leis de acesso aberto à produção científica. Com isso, elas promovem a ampliação do acesso aberto em várias regiões e fomentam a conscientização no âmbito político e de pesquisa quanto à relevância de articulação entre os diversos setores partícipes na cadeia de produção e divulgação científica.

Nesta perspectiva, foi criada a Rede Federada de Repositórios Institucionais de Publicações Científicas (LA Referencia), uma rede composta por nove países da América Latina (Argentina, Brasil, Chile, Colômbia, Equador, El Salvador, México, Peru e Venezuela) cujo objetivo é estabelecer uma estratégia comum em relação à interoperabilidade e à gestão dos repositórios institucionais de publicações científicas da região. Ela também oferece serviço de busca integrada em sistemas de informação dos países integrantes, configurando-se como um relevante serviço de agregação de repositórios.

Desde a sua implementação, em 2012, a rede vem promovendo o acesso aberto aos resultados de pesquisa financiadas com recursos públicos com o intuito de dar visibilidade à produção científica das instituições de educação superior e pesquisa da América Latina. Chan, Kirsop e Arunachalam (2005), elucidam que a constituição de um conjunto de repositórios em acesso aberto possibilita à comunidade científica regional ter mais visibilidade na comunidade científica internacional, com isso valoriza-se a produção local de pesquisa e conhecimento. Estas ações podem contribuir para a resolução de problemas mundiais em saúde, agricultura, meio ambiente e demais áreas específicas.

Costa e Leite (2013) destacam que as iniciativas de acesso aberto na América Latina têm sido importantes para superar as limitações enraizadas tanto no acesso à literatura internacional quanto na disseminação do conhecimento científico produzido na região. Diversas ações têm sido promovidas pelas redes, uma delas é o desenvolvimento de parcerias com redes globais de repositórios como a *Confederation of Open Access Repositories* (COAR) que incentiva colaboração internacional e a criação de uma rede global de cooperação de repositórios digitais.

Nesse sentido, o documento “Políticas para a ciência aberta e os dados científicos na América Latina”, produzido pelo Comitê Diretivo da LA Referencia, ressalta o caráter não comercial da produção científica na América Latina, característica atribuída pelo majoritário financiamento das pesquisas pelo Estado, de maneira direta e indireta. Esclarece que o volume de produções em repositórios e revistas científicas de acesso aberto é o dobro daquelas publicadas em bases de dados de acesso fechado (LA Referencia, 2018). Assim, a proposta de desenvolvimento de redes como a LA Referencia deve estar pautada na realidade regional, considerando o uso de padrões e tecnologias abertas com abordagem não comercial, visando à construção de plataformas públicas que se conectem às demais iniciativas internacionais, possibilitando a interoperabilidade, recuperação e divulgação das produções incluídas.

Tendo em vista que a despeito da expressividade da produção científica realizada por países periféricos sua divulgação e visibilidade internacionais ainda tem alcance reduzido, o objetivo deste trabalho é analisar as ações desenvolvidas pela Rede LA Referência relacionadas à ampliação do acesso aberto na América Latina a fim de conhecer e explicitar os elementos que subsidiam a formação de redes de repositórios.

A proposta se justifica pela relevância de explorar iniciativas que, estando fora do eixo central da produção científica, possam se tornar referência na formação de novas redes de repositórios que almejem estabelecer políticas comuns e ampliar sua visibilidade em nível regional ou mundial. Considerando a pesquisa de Alperim (2016), o modelo regional para acesso aberto instituído na América Latina apoia o conhecimento enquanto bem público na medida em que o coloca em disponibilidade ampla de acesso comum e com um alcance sem precedentes que ultrapassa o âmbito acadêmico. Ações como a criação de uma base tecnológica que fornece recursos de coleta, desenvolvimento colaborativo e transferência de soluções para nós nacionais; acordos de padrões de metadados essenciais para recuperação da informação; infraestrutura científica com tecnologia interoperável; além de diretrizes e desenvolvimento de sistemas que permitem que a rede cumpra com os objetivos de visibilidade e qualidade da informação, fazem da LA Referência uma experiência exitosa na região.

Para Casal Reyes (2018) iniciativas como esta cumprem relevante papel pois integram repositórios em serviços de consulta mais amplos, configurando-se como meio privilegiado para que os pesquisadores tenham acesso a conteúdos disponíveis em acesso aberto, bem como possam atestar a validade de optar por esta forma de publicação. Ademais, elas multiplicam os caminhos de acesso que permitem aceder a informações provenientes de diversas fontes e produtores, especialmente em círculos não acadêmicos. Neste sentido, faz-se pertinente destacar as diretrizes e ações realizadas pela rede de modo a sublinhar elementos que sustentam a criação de uma política comum ampla e conceituada, e que poderão nortear ações de mesmo relevo ou motivar o engajamento de novos membros.

Procedimentos metodológicos

A pesquisa desenvolvida é qualitativa, de caráter descritivo e utiliza como técnicas de investigação a revisão de literatura e a análise documental de acordos políticos e técnicos, além de diretrizes estabelecidas pela Rede LA Referência e deliberações do Conselho Diretivo.

Para identificar o histórico de constituição da Rede e a estruturação de suas ações serão analisados os acordos políticos e técnicos, as apresentações realizadas pelos membros do Conselho Diretivo, as diretrizes propostas com relação aos critérios de qualidade da informação e estruturação de metadados, as estratégias de coleta dos dados e as regras estabelecidas. Além disso, serão analisados os aspectos que levaram à definição de interoperabilidade da LA Referência com as organizações *Open Access Infrastructure for Research*

in Europe (OpenAIRE) e *Confederation of Open Access Repositories* (COAR), tendo em vista o objetivo comum de desenvolvimento de plataformas internacionais de colaboração em repositórios de acesso aberto para apoiar pesquisas e comunicações científicas.

A seguir, são apresentados aspectos iniciais a serem analisados no presente artigo.

Principais ações de acesso aberto desenvolvidas pela Rede Federada de Repositórios Institucionais de Publicações Científicas

Ao longo do seu desenvolvimento a LA Referencia atuou na constituição de ações de divulgação e fortalecimento do acesso aberto e ciência aberta, além de construir alianças com outras redes para o compartilhamento e visibilidade da produção gerada pelas instituições de ensino e pesquisa da América Latina e Caribe. Apresenta-se, no Quadro 1, algumas das principais deliberações da LA Referencia, com relação à sua constituição e desenvolvimento.

Quadro 1 - Principais deliberações da Rede LA Referencia

Data	Ação	Objetivos
2011	Declaração de Acesso Aberto para a Rede Federada Latinoamericana de Repositórios Institucionais de Documentação Científica	<p>A Rede adere à Declaração de Berlim. Como manifestação de tal compromisso e em favor de seu desenvolvimento, os membros promoverão:</p> <ul style="list-style-type: none"> - Que pesquisadores e beneficiários de fundos públicos publiquem os resultados da pesquisa, de acordo com os princípios de acesso público; - O desenvolvimento de ferramentas e mecanismos para avaliar as contribuições em acesso aberto e periódicos on-line, a fim de manter padrões de garantia de qualidade e boas práticas científicas; - Gerar ferramentas que permitam medir a produção científica de repositórios da região, com padrões de garantia de qualidade e baseados em boas práticas científicas.
2011	Acordo de Promoção de Políticas e Incentivos de Acesso Aberto	<ul style="list-style-type: none"> - Criação de um espaço dentro da plataforma da Rede para a promoção do acesso aberto que terá como objetivo tornar visível e documentar pelos membros da Rede as iniciativas da região que são exitosas na promoção do acesso aberto em todos os níveis e gerar e/ou divulgar instrumentos para fomentar a apropriação do acesso aberto pelos principais atores envolvidos no processo de comunicação científica (tomadores de decisão, autores, editores, serviços de informação, etc); - Impulsionar políticas para promover o depósito de resultados de investigação em repositórios de acesso aberto que implementem o metadado "rights" do Dublin Core qualificado; - Gerar estratégias de divulgação e socialização que permitam o reconhecimento da Rede na comunidade

acadêmica e científica da região.

- | | | |
|------|--|---|
| 2012 | Acordo de Cooperação entre autoridades de ciência, tecnologia e inovação da América Latina para constituição da LA Referencia. | Criação e desenho da Rede Federada Latino-Americana de Repositórios de Documentação Científica (LA Referencia), cuja missão é "promover a cooperação entre os países membros para acesso, recuperação e visibilidade da produção científica na América Latina, como bem público regional". |
| 2015 | Projeto de internacionalização: interoperabilidade com OpenAIRE e COAR | LA Referencia e OpenAIRE passam a adotar estruturas comuns de metadados, baseadas nas atuais diretrizes do OpenAIRE e nos vocabulários que estão sendo desenvolvidos no contexto do COAR, para as redes de repositórios, desenvolvendo as diretrizes em colaboração; COAR, LA Referencia e OpenAIRE visam identificar formas de gerenciamento de repositórios e redes de repositórios na América Latina, particularmente em termos da implementação de melhores práticas; A partir das indicações do COAR, LA Referencia e OpenAIRE continuarão a trabalhar com outras redes nacionais e regionais para promover a visão de uma rede global de repositórios e investigar serviços comuns de apoio ao acesso aberto e ciência aberta |
| 2018 | Publicação do documento: Políticas para a ciência aberta e dados científicos na América Latina | <ul style="list-style-type: none"> - Apoio a dados científicos e validação de publicações que tratam da reprodutibilidade da ciência, coleta e geração de dados de pesquisas financiadas com recursos públicos e que lidam com planos de gerenciamento de dados; - Busca de soluções para questões relacionadas a softwares e ferramentas necessárias para entender e reutilizar os dados de pesquisa, além de metadados interoperáveis padrão; - Cooperação com as iniciativas que visam encorajar e propor planos de gestão de dados nas instituições que os produzem. |

Fonte: LA Referencia, 2019, adaptado pelas autoras.

A partir das iniciativas elencadas pode-se identificar os esforços dos constituintes da Rede de garantir uma integração entre as iniciativas de acesso aberto e ciência aberta constantes nos países integrantes por meio da padronização dos dados e definição de estratégias comuns que visam aumentar a visibilidade e a recuperação da produção científica da região.

Considerações e reflexões

O acesso aberto promovido pelas redes regionais integradoras de repositórios contribui para a economia de recursos, amplia a capacidade de atuação para divulgação da produção científica em âmbito internacional,

aglutina investimentos e força político-econômica. Além disso, contribui para diminuir as desigualdades nacionais e regionais de acesso aberto à produção científica, promove visibilidade dos autores, das universidades, das instituições de pesquisas e das agências de fomento para um público mais amplo do que o acadêmico.

A Rede LA Referencia, instituída por países da América Latina com o objetivo de compartilhar, fomentar visibilidade e acesso igualitário à produção científica da região, tem como propósito principal a promoção do acesso aberto, incentivando o uso de licenças que facilitem o uso e a reutilização dos documentos e dados de pesquisa. Trata-se de um relevante exemplo de rede que se desenvolve com êxito, por meio da definição de políticas, tecnologias e estratégias de divulgação e socialização, tendo alcançado reconhecimento de sua relevância perante o cenário científico internacional.

Referências

ALPERIM, Juan Pablo. **The Public Impact of Latin America's Approach to Open Access**. 2015. 145p. Dissertation (Doctorship in Philosophy) - Stanford University, Stanford, 2015. Disponível em:

<https://stacks.stanford.edu/file/druid:jr256tk1194/AlperinDissertationFinalPublicImpact-augmented.pdf>. Acesso em: 22 abr. 2019.

CASAL REYES, María Isabel. Repositorios, conceptos principales y elementos para su gestión. In: GONZÁLEZ, José Diego (Dir.). **Acceso abierto | Dossier Cerlalc | Derecho de autor**. Bogotá: Cerlalc, 2018. p. 73-85. Disponível em: <https://universoabierto.org/2018/03/19/acceso-abierto-dossier-cerlalc-derecho-de-autor>. Acesso em 22 abr. 2019.

CHAN, L; COSTA, S. Participation in the global knowledge commons: challenges and opportunities for research dissemination in developing countries. **New Library World**, v. 106, n. 3/4, p.141–163. 2005. Disponível em: 10.1108/03074800510587354 Acesso em: 22 abr. 2019.

CHAN, Leslie; KIRSOP, Barbara; ARUNACHALAM, Subbiah. **Open access archiving**: the fast track to building research capacity in developing countries, 2005. Disponível em: https://tspace.library.utoronto.ca/bitstream/1807/4415/1/Open_Access_Archiving.pdf. Acesso em: 22 abr. 2019.

LA REFERENCIA. Red de Repositorios de Acceso Abierto a la Ciencia. [S. l.], 2019. Disponível em: <http://www.lareferencia.info/pt/>. Acesso em: 22 abr. 2019.

LA REFERENCIA. Red de Repositorios de Acceso Abierto a la Ciencia. Políticas para a ciência aberta e os dados científicos na América Latina. 2018. Disponível em: <http://www.lareferencia.info/pt/recursos/ciencia-abierta-documentos-externos>. Acesso em: 22 abr. 2019.

